DOCUMENT RESUME

ED 298 286 CE 050 799

AUTHOR	Friedlander, Daniel
TITLE	Subgroup Impacts and Performance Indicators for Selected Welfare Employment Programs.
INSTITUTION	Manpower Demonstration Research Corp., New York, N.Y.
SPONS AGENCY	Department of Health and Human Services, Washington, D.C.; National Commission for Employment Policy (DOL), Washington, D.C.
PUB DATE	Aug 88
NOTE	184p.
AVAILABLE FROM	Manpower Demonstration Research Corporation, Three Park Avenue, New York, NY 10016.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	*Admission Criteria; Comparative Analysis; Employment Patterns; *Employment Programs; *Outcomes of Education; *Performance; Policy Formation; *Program Effectiveness; Salary Wage Differentials; Selective Admission; *Welfare Recipients
IDENTIFIERS	Aid to Families with Dependent Children; Work Incentive Program

ABSTRACT

A study examined the effectiveness of five mandatory welfare employment programs in working with different segments of the Aid to Families with Dependent Children (AFDC) caseload. The study considered programs in San Diego, California; Baltimore, Maryland; several counties in Virginia; Little Rock and another county in Arkansas; and Cook County (including Chicago) in Illinois. The experimental groups consisted of persons who could receive the special service programs. Welfare recipients assigned to the control groups could not receive the services under study. They could, however, avail themselves of other community services (on their own initiative). The impacts and outcomes identified indicate that, if maximizing program impacts while limiting resources is the goal, it is a mistake to concentrate only on serving the most job-ready portion of the AFDC population. It also appeared relatively clear that programs should not focus exclusively on the most disadvantaged among the Work Incentive (WIN)-mandatory caseload (at least not with low- to moderate-cost services). The data collected suggest that program operators wishing to maximize impacts on earnings may find it desirable to work first with applicant returnees or applicants with weak work records, whereas those operators desiring to maximize welfare savings devote increased effort to the most dependent client groups. (MN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

BOARD OF DIRECTORS

RICHARD P. NATHAN, *Chairman*
Professor
Woodrow Wilson School of
Public and International Affairs
Princeton University

M. CARL HOLMAN, *Vice-Chairman*
President
National Urban Coalition

PAUL H. O'NEILL, *Treasurer*
Chairman and CEO
Aluminum Company of America

ELI GINZBERG, *Chairman Emeritus*
Director
Conservation of Human Resources
Columbia University

BERNARD E. ANDERSON
Managing Partner
Urban Affairs Partnership Corporation

RAMON C. CORTINES
Superintendent
San Francisco Unified School District

ALAN KISTLER
President
Human Resource Development Institute
AFL-CIO

ELEANOR HOLMES NORTON
Professor
Georgetown University Law Center

ISABEL V. SAWHILL
Senior Fellow
The Urban Institute

ROBERT SOLOW
Institute Professor
Massachusetts Institute of Technology

GILBERT STEINER
Senior Fellow
Brookings Institution

MITCHELL SVIRIDOFF
Professor of Urban Policy
New School for Social Research

WILLIAM S. WOODSIDE
Chairman
Executive Committee of the Board of Directors
Primerica Corporation

JUDITH M. GUERON
President
Manpower Demonstration Research Corporation

MDRC

SUBGROUP IMPACTS AND PERFORMANCE INDICATORS FOR SELECTED WELFARE EMPLOYMENT PROGRAMS

Daniel Friedlander

Manpower Demonstration
Research Corporation

August 1988

BEST COPY AVAILABLE

This analysis completes the second phase of research conducted on subgroup impacts and performance measurement in welfare employment programs. The research was funded by the Office of the Assistant Secretary for Planning and Evaluation at the U.S. Department of Health and Human Services; by the Office of Family Assistance, Family Support Administration, also part of the U.S. Department of Health and Human Services; and by the National Commission for Employment Policy. The findings and conclusions of this report do not necessarily represent the official position or policies of the funders.

Copyright 1988 Manpower Demonstration Research Corporation

ACKNOWLEDGMENTS

The author wishes to thank those at the U.S. Department of Health and Human Services (the Office of Assistant Secretary for Planning and Evaluation and the Family Support Administration) and at the National Commission for Employment Policy who made this research possible. Several staff members from these agencies also participated in reviews of the findings and provided thoughtful comments throughout the analysis. In particular, the author is grateful for the ongoing input and suggestions from Kay Albright, Deirdre Duzor, Canta Pian, Howard Rolston, Steve Sandell and Daniel Weinberg. Valuable comments at a project briefing were also offered by Robert Helms, Wayne Stanton, Carl Dahlman and Winifred Smith. In addition, the contribution of Claudia Kirk, Ken Lee, Gay Morris and M.I. Pendell was appreciated.

The study benefited greatly from the suggestions and advice of Gary Burtless at the Brookings Institution, David Ellwood of Harvard University and Frank Levy of the University of Maryland. Gratitude is also expressed to the Work/Welfare Committee of MDRC's Board of Directors and to Gordon Berlin of the Ford Foundation who reviewed and commented on the results of this research as did the special Advisory Committee to the Work/Welfare Demonstration.

Because this study is based on data collected in MDRC's ongoing Demonstration of State Work/Welfare Initiatives, the author also expresses gratitude to many members of MDRC staff who have obtained the data and produced the state reports over the last few years. Several of the analyses discussed here stem from an earlier report for this project that was co-authored with David Long, who set up the benfit-cost analyses for the Work/Welfare Demonstration. Other MDRC staff who were also instrumental in shaping this report include Judith Gueron, Michael Borus, Barbara Goldman, David Long and John Wallace. In this connection, special thanks go to Michael Bangser, whose insightful critiques provided invaluable feedback at critical stages in the research. Another debt is owed to Felicity Skidmore for her skillful and rigorous editing of the manuscript.

Additionally, at MDRC, David Jaeger and Johanna Walter prepared the data sets and carried out the computer analysis, building on work begun by Jan Bryant and Marjorie Erickson. Ginger Knox and Emma Caspar worked on benefit-cost calculations. Stephanie Powell and Rudd Kierstead provided additional programming support. Copyediting was ably done by Alex Kopelman. Celeste Fuller prepared and painstakingly edited the tables. Patt Pontevolpe, Claudette Edwards and Stephanie Amy Cowell efficiently handled the word processing.

The Author

PREFACE

This is a special report generated by research from MDRC's Demonstration of State Work/Welfare Initiatives. This demonstration is a unique opportunity for MDRC to work with states in evaluating their employment programs and thus to examine the potential effectiveness of a major component of recent welfare reform proposals.

Using data from five state welfare employment programs (those in San Diego, Baltimore, Virginia, Arkansas, and Cook County, Illinois), the study presented here has two purposes, both of which are important in designing and operating effective programs. One is to produce reliable estimates of the programs' relative impacts on the employment and welfare receipt of different groups of welfare applicants and recipients. The second objective is to help develop valid operational indicators for measuring the success of different welfare employment programs.

The search for reliable and workable standards of performance to be used in employment programs for welfare recipients is one of the major themes in current efforts at welfare reform. MDRC hopes that these findings will contribute to informed decision-making on this subject and ultimately to the development and operation of even more effecitve programs designed to increase the self-sufficiency of all welfare recipients.

Judith M. Gueron
President

EXECUTIVE SUMMARY

This report presents an analysis of the effectiveness of five mandatory welfare employment programs in working with different segments of the Aid to Families with Dependent Children (AFDC) caseload. Among their several goals, these programs all sought to increase earnings and decrease dependence on welfare, although local planners assigned different relative importance to one or the other objective.

The analysis here has two purposes, both of which are useful in designing and operating programs. One is to produce estimates of the programs' relative impacts on the employment and welfare receipt of different groups of welfare applicants and recipients. These estimates may provide useful information to guide the targeting choices of policymakers who wish to maximize program impacts with limited budgets. The other objective is to explore the validity of certain principles of performance measurement in an effort to assist in the development of operational indicators that will best encourage the long-term objectives of maximizing earnings gains and reductions in welfare dependency.

The Programs Evaluated

The analysis is based on data collected in evaluations of welfare employment programs in San Diego, Baltimore, several counties in Virginia, Little Rock and one other county in Arkansas, and Cook County (containing the City of Chicago) in Illinois. These programs required the participation of portions of the AFDC caseload which are "mandatory" under federal

8

Work Incentive (WIN) Program regulations (i.e., primarily women whose youngest child is six years old or older). The programs provided different services and operated in different labor markets, but all relied primarily on a combination of job search and work experience. Program costs (in 1987 dollars) ranged from a low of $150 per experimental in Cook County to a high of $1,050 in Baltimore. AFDC income eligibility regulations also varied, with the highest benefit standards in California and the lowest in Arkansas.

San Diego enrolled all WIN-mandatory welfare applicants but did not enroll persons who were already AFDC recipients. Participants went through a three-week job search workshop, followed by a 13-week work experience obligation for those who had not found an unsubsidized job. Baltimore enrolled WIN-mandatory applicants and persons who were recipients but had just become mandatory, usually because their youngest child had turned six years of age. Program activities could be selected from a number of job search, work experience, education and training options. The Baltimore program restricted active enrollment to 1,000 registrants per year during the period studied.

Virginia enrolled its entire WIN-mandatory caseload. Job search was required of all enrollees and was followed, at county option, by short-term work experience, education or training. Arkansas enrolled its entire WIN-mandatory caseload, but only applicants and recipients who became mandatory after the research began were included in the impact sample. The program consisted primarily of independent and group job search and, less frequently, a work experience component. Arkansas also obtained a waiver enabling it to classify as WIN-mandatory AFDC mothers whose youngest child was three to five years of age.

Cook County, with one of the nation's largest urban caseloads, worked with recipients and the subset of applicants whose grants had been approved. As with Arkansas and Baltimore, the Cook County impact sample was restricted to those who became WIN-mandatory after the research began. The Cook County program included a two-month job search reporting component that relied mostly on the initiative of clients backed up by routine sanctioning for noncompliance. A private nonprofit work experience position was assigned at a later date for many of the individuals who did not find employment through the job search component. Work experience in all these programs typically lasted not more than three months; in no case was it designed to continue for as long as participants remained on AFDC.

The Research Design

All five evaluations used research designs in which eligible applicants and recipients were randomly assigned to experimental groups, which could receive the special program services, or to control groups, which could not. The experience of the control group members -- who, on their own initiative, were able to avail themselves of services elsewhere in the community -- indicates what would have happened to the experimental groups in the absence of the special intervention, affording a benchmark for measuring program impacts.

It should be noted that this study labels as an applicant any person who was in the process of applying for welfare at the time of random assignment. Applicants retained the applicant designation even if they were approved and began receiving welfare; even if they left welfare; and even if their application was never approved. A recipient is anyone who was already receiving AFDC at the time of random assignment. Recipients retained the recipient label even if they left welfare.

The data on which the analysis is based were collected from Unemployment Insurance earnings records and automated AFDC payment ledgers for varying follow-up periods: a minimum of three years in Arkansas, two and a half in Baltimore and Virginia, and a year and a half in San Diego and Cook County. Average quarterly impacts are estimated on the basis of data from the fourth quarter after enrollment through the end of the follow-up period. The subgroup analysis focused on heads of single-parent households (primarily women). Two-parent households (mostly headed by men eligible under the AFDC-Unemployed Parent program) were included in two of the individual program evaluations but are not included in this study.

Impact estimates are reported on a per-experimental basis, even though only about half the experimentals on average actually participated in some formal activity. The estimates therefore represent the program effect averaged over all program registrants, not just participants. The estimates of average earnings and average AFDC payments also include all sample members, counting as zeroes those who did not work or did not receive welfare. Some special statistical considerations relevant to comparisons of subgroups are laid out in an appendix to the report.

The analysis first examines impacts across subgroups that vary in their prior employment, welfare history and other demographic characteristics. It then uses subgroup impacts to evaluate two frequently used performance measures -- the number of "job entries" (placements) and the number of cases "off-welfare" (case closures).

The Distinction Between Outcomes and Impacts

The distinction between the meanings of "outcomes" and "impacts" as defined for this analysis is critical to understanding the findings. An outcome is the employment or welfare status of a person at a specified point after program enrollment. An impact is the change in an outcome produced by a program during that period. Program impact is estimated as the difference in outcomes between the experimental and control groups.

Program impacts are smaller than outcomes because the normal job-finding and welfare-departure rates of the AFDC population -- i.e., control group outcomes -- are not zero in the absence of a new program. But the relative difference between outcomes and impacts is not the same across all subgroups. Some subgroups exhibit worse-than-average outcomes but generate better-than-average impacts; other subgroups do the reverse.

For example, in San Diego, experimental applicants with $3,000 or more earnings in the year before enrollment attained an average quarterly employment rate of 61 percent during the second year of follow-up. This is a high rate compared to other subgroups in these samples of AFDC mothers. But controls with the same prior earnings did almost as well, even without the special intervention. They averaged a 59 percent employment rate over the same follow-up period. The increase -- i.e., the impact of the program -- for these individuals was therefore only 2 percentage points. Clearly, the high outcome levels of employment reported for this group grossly overstate the influence of the program. In contrast, experimental applicants who had no prior earnings attained only a 30 percent employment rate, less than half that of the "more employable" group. This outcome

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

---

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

S E Mandel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

level, however, amounted to nearly an 8 percentage point increase over comparable controls. Thus, although the employment outcomes for this group seem on the face of it to be worse, their impacts are, in fact, larger. Analogous examples could be given for welfare outcomes.

If the example discussed is not unique -- and the research reported in the next section indicates that it is not -- then policymakers are faced with a serious dilemma. On the one hand, they may deem it important to impress upon local operators that the ultimate program goals are employment and departure from welfare. On the other hand, by encouraging programs to strive for high rates of "placement" or "job entry" and high rates of welfare case closure, they may be driving operators to focus attention on groups of clients for whom impacts are below average. Thus, standards of performance based on simple outcomes, at best, may be unrelated to real program performance and, at worst, may even tend to undermine true effectiveness.

The Major Subgroups

Samples for the five programs were divided into a variety of subgroups, with impacts on earnings and welfare receipt estimated for each. The main objective of the analysis was to focus on subgroup definitions that might be of practical use in targeting program services. These subgroup definitions had to meet four criteria: (1) special targeting to the subgroup would not automatically be ruled out on political grounds; (2) the required background information is objective and verifiable; (3) the required background information can be obtained cheaply at program enrollment; and (4) the subgroup constitutes a meaningful share of the eligible caseload. Prior earnings and welfare history turned out to be the subgroup dimensions that met these criteria and best predicted future employment and welfare receipt -- the two outcomes of greatest interest for welfare employment programs. Several other individual characteristics were also investigated, including some which do not meet the four criteria for practical application in targeting but which are of interest nonetheless.

The principal subgroup division used was by applicant/recipient status, with applicants further divided into first-time applicants and applicant returnees (i.e., applicants who had prior welfare experience, have gone off welfare, and have returned to welfare for some reason). Within the applicant and recipient categories, three major subgroups were defined based on earnings from employment in the year prior to random assignment: no earnings, earnings of $1 to $2,999, and earnings of $3,000 or more. Three other major subgroups were created according to length of time prior to random assignment that the sample member had had her own AFDC case: never, two years or less, and more than two years.

Obviously these characteristics can be combined in different ways to produce a variety of subgroup configurations. One particularly promising configuration is shown in Tables 1 and 2, grouped in three mutually exclusive tiers and arranged in roughly ascending order of welfare dependence and descending order of employability. The least dependent group, for example, is new applicants with no prior AFDC. The subgroups within a tier overlap, constituting alternative ways of grouping individuals. Depending on location, the first tier comprised from 25 to over 55 percent of the applicant sample, the second tier between 45 and 75 percent. For the four programs that enrolled recipients, the entire third

TABLE 1

SUMMARY OF IMPACTS ON QUARTERLY EARNINGS FOR MAJOR SUBGROUPS OF AFDC APPLICANTS AND RECIPIENTS

Subgroup	San Diego	Baltimore	Virginia	Arkansas	Cook County[a]
First Tier					
Applicants with No Prior AFDC	$ +37	$ +121	$ -13	$ +26	$ ---
Second Tier					
Applicant Returnees	+158**	+188***	+114*	+211***	---
Applicant Returnees with Less than $3000 Prior Earnings	+151**	+253***	+20	+202**	---
Third Tier					
All Recipients	---	+37	+69*	+19	+46**
Recipients with More than Two Years on AFDC	---	-0	+110**	+14	---
Recipients with No Prior Earnings	---	+104**	+70	+29	+12
Recipients with No Prior Earnings and More than Two Years on AFDC	---	+88	+94*	+28	---
All AFDC					
Quarterly Earnings Impact	+118**	+96***	+72**	+70**	+19
Average Control-Group Earnings	773	634	541	257	451

NOTES: Tiers are mutually exclusive; subgroups within tiers overlap. All values are averages for the fourth through the last quarter of follow-up. Estimates include zero values for sample members not employed or for sample members not receiving welfare.

A two-tailed t-test was applied to differences between experimental and control groups. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent, *** = 1 percent.

[a] The definitions of "applicant" and "recipient" for Cook County are not strictly comparable to those of the other programs.

TABLE 2

SUMMARY OF IMPACTS ON QUARTERLY AFDC PAYMENTS FOR MAJOR SUBGROUPS OF AFDC APPLICANTS AND RECIPIENTS

Subgroup	San Diego	Baltimore	Virginia	Arkansas	Cook County[c]
First Tier					
Applicants with No Prior AFDC	$ -5	$ -9	$ -28	$ -31	$ ---
Second Tier					
Applicant Returnees	-47	-15	-16	-19	---
Applicant Returnees with Less than $3000 Prior Earnings	-63*	-19	-29	-22	---
Third Tier					
All Recipients	---	+5	-24	-60***	-13
Recipients with More than Two Years on AFDC	---	+19	-48**	-44*	---
Recipients with No Prior Earnings	---	+1	-26	-63***	-6
Recipients with No Prior Earnings and More than Two Years on AFDC	---	-1	-48**	-48*	---
All AFDC					
Quarterly AFDC Payments Impact	-33	-5	-23*	-40***	-13
Average Control-Group AFDC Payments	469	501	345	232	646

NOTES: Tiers are mutually exclusive; subgroups within tiers overlap. All values are averages for the fourth through the last quarter of follow-up. Estimates include zero values for sample members not employed or for sample members not receiving welfare.

A two-tailed t-test was applied to differences between experimental and control groups. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent, *** = 1 percent.

[a] The definitions of "applicant" and "recipient" for Cook County are not strictly comparable to those of the other programs.

tier accounted for about 40 to over 65 percent of the full sample.

Some idea of the relative employability and welfare dependence of the different tiers is given by the experience of control group members -- specifically, their employment in quarters 4 to 6 after enrollment and their welfare receipt in quarter 6. For the first tier group, for example, 37 percent of controls were employed and 31 percent were receiving welfare. For the second tier groups, 28 to 37 percent of the control groups were employed and 46 to 49 percent were receiving welfare. For the third tier subgroups, 13 to 23 percent were employed, and 72 to 80 percent were on welfare.

In assessing the findings, attention should be paid to the magnitude of impacts as well as their statistical significance. One way to do this is to compare a subgroup impact to the approximate median value of all subgroup impacts in the five samples, i.e., the value for which about half the impact estimates fall above and half below. Statistically significant impacts above this overall average (roughly $100 per quarter for earnings increases and $20 per quarter for welfare savings) are called "above average." Impacts below this level are called "below average."

In interpreting the results, it should be born in mind that (a) the samples consist primarily of adult women without pre-school children and (b) the interventions observed were mandatory, mass participation programs, with a low-to-moderate cost per enrollee. The findings may not be directly generalizable to other poverty groups or to other kinds of programs. For example, information about a client's past earnings and AFDC receipt may have less significance for youth, who typically have only short work and welfare histories.

Impacts for Subgroups Defined by Prior Employment and Welfare History

- **The groups that were most job-ready and least welfare-dependent, as defined by previous work and welfare experience, had below-average program impacts that were generally not statistically significant.**

Impacts for the first tier subgroups were generally low. Applicants with no prior AFDC history -- first-time applicants -- had quarterly earnings gains over $100 per quarter in only one sample, and these were not statistically significant. (See Table 1.) Welfare savings were also small for this group and were never statistically significant. (See Table 2.)

Another subgroup (not presented in Tables 1 and 2) which might be considered in the first tier is applicants with $3,000 or more in prior-year earnings. This group showed significant earnings gains in only one sample and never showed welfare savings. Additional analysis of the one location with earnings gains revealed that nearly all of the increase accrued to individuals who would have left welfare even without the special intervention. This suggests that, even when the least dependent do achieve earnings gains, these may not translate into reductions in AFDC receipt.

- **Earnings impacts were found most consistently for individuals in the mid-dependency tier.**

The second tier contains applicants with some prior welfare history (i.e., returnees). Also shown separately in Tables 1 and 2 is the subset of these returnees who earned less than $3,000 in the year prior to their application for welfare. Compared to the other two tiers, these groups proved more likely to have above average and statistically significant earnings impacts, although this was not true in all cases. Moreover, additional statistical tests showed that earnings impacts for applicant returnees were significantly greater than those for new applicants and recipients combined in two out of the three cases where such comparisons were possible.

- **Earnings impacts were not found consistently for subgroups in the most dependent tier.**

The third tier contains the full sample of welfare recipients in each of the four locations where they were enrolled, with results also presented for several overlapping subgroups of recipients. Employment and earnings impacts were found for some of these subgroups, including some gains that were statistically significant. However, these were not consistent and were rarely above average. This was not a result, however, of the programs' failure to serve this population. On the contrary, participation rates for these groups were as high as, or higher than, the rates for applicants.

- **Statistically significant welfare savings were found for several of the more dependent subgroups, but this was not consistent across samples. Welfare savings were also found for some of the mid-dependency groups, although generally these were not statistically significant.**

Welfare savings, because they were smaller than earnings increases, were more difficult to contrast across subgroups (see Table 2). Modest savings were found for subgroups in the second tier. The individual estimates were usually not statistically significant. Interestingly, in two samples (Virginia and Arkansas), some of the more dependent subgroups showed welfare reductions that were relatively large, especially in comparison to their earnings gains.

- **The findings do not support a strong recommendation to narrowly target low-to-moderate cost services. Where increasing earnings is the primary program objective, subgroups in the mid-dependency tier may be the best candidates for priority**

> **attention. Where program objectives emphasize reductions in welfare payments, groups in the most dependent tier may assume increased importance.**

No subgroup emerges clearly and consistently as appropriate for exclusive targeting of low-to-moderate cost services. While not conclusive, the findings suggest, moreover, that selection of a targeting strategy might depend on the primacy of different potential program objectives. When resources are scarce, administrators seeking to increase the earnings of those on welfare might assign priority for services to subgroups in the second tier. These groups -- applicant returnees -- showed the most consistent earnings gains. Alternatively, administrators seeking welfare savings should recognize that a substantial share of estimated savings were found in the high-dependency third tier.

The particular programs tested, however, suggest caution in using these findings to reach a final conclusion on targeting. The results come from low-to-moderate cost programs that explicitly sought to include the full range of individuals within the groups served. Inclusion of the more job-ready, tier-one registrants might not be suggested by the impact data, yet this group may have played an important role by providing encouragement to the less job-ready participants and to program staff. Further, administrators have a choice not only in who they target but in the services they provide. For recipients in tier three, low-to-moderate cost programs do not consistently produce above-average earnings gains. It remains unclear, however, whether more intensive or different services could lead to more consistent earnings impacts, while also producing welfare savings.

Other Subgroup Impacts

In addition to prior work and welfare history, subgroups were defined along several demographic dimensions, including marital status, education, and the number and ages of children. These factors, although important, were less strongly related to future employment and welfare receipt.

- **The samples yielded often conflicting impact estimates for subgroups of the WIN-mandatory AFDC caseload defined on characteristics other than prior employment and welfare experience.**

High school diploma status, absence of children under age twelve, number of children, age and ethnicity did not produce consistent impact differentials across program samples. Among sample members without a diploma, earnings impacts were found only for applicants and then only in Baltimore and Virginia. The education and training services available in these two programs may have played a role in achieving results for dropouts there, but this is by no means a clear-cut conclusion, particularly since controls in Virginia were able to obtain similar services on their own at about the same rate as experimentals.

There was some evidence that the bulk of welfare savings were obtained for women who were not married at the date of random assignment, particularly those who were never married. Evidence also suggests that impacts among recipients in rural areas were weak, and that this may account for some part of the differences between applicants and recipients overall.

- **In Arkansas, mandatory status was extended to women whose youngest child was three to five years old. Impacts for these women were about the same as impacts for the regular WIN-mandatory enrollees.**

The inclusion of women with children ages three to five in Arkansas more than doubled the number of individuals who enrolled in the program during the demonstration. Employment rates were the same for this group as for women with older children. Program impacts on earnings and welfare receipt were also similar. The total effects of the program on the AFDC caseload were therefore more than twice what they would have been if only the impacts on regular WIN mandatories were counted. Some caution should be exercised in generalizing this finding to other program contexts, however, since AFDC grant levels are low in Arkansas and the sample size was relatively small.

- **The combination of impact differences across dependency subgroups and according to demographic characteristics suggests a possible threshold effect for earnings impacts.**

For cases whose multiple disadvantages combine to make them more dependent than some threshold level, the typically low-cost services provided by the programs in this study may begin to lose their effectiveness in increasing earnings. It is not clear how large the group below the threshold may be, or whether it can be adequately identified with demographic data alone. It was found, however, that sample members who were recipients with more than two years on welfare, without recent earnings, and with no high school diploma attained below-average earnings impacts in the three samples for which data on these characteristics were available.

Program Performance Measures

In this study, as noted, a program's performance is defined as the impact it achieves. Normally, administrators do not have impact measures available to them. Instead, measures of outcomes or participation must be utilized to set operational goals and standards of performance. The most popular of these performance indicators have been "job entries" (placements) and cases "off welfare" (case closures). For programs that seek to maximize impacts, these performance indicators can only be effective if they are related to impacts.

The validity of performance measures based on observable employment and welfare outcomes was assessed by examining the correlation between job entries and case closures and program impacts. Unemployment Insurance earnings were used to identify a job entry; AFDC payments were used to determine whether a sample member went off welfare.

- **Unadjusted job-entry and off-welfare measures are not empirically valid indicators of performance if program objectives are to increase earnings and decrease welfare receipt.**

No consistent relationship emerged across program samples between simple outcome indicators (i.e., job entries and case closures) and program impacts. In San Diego and Baltimore, in fact, subgroups with the higher job-entry scores had lower earnings and welfare impacts. In Virginia and Arkansas, job entries indicated employment impacts mostly for sample members with relatively low risk of remaining on welfare a long time. Job entries were therefore negatively related to impacts on welfare receipt.

The study identified two major reasons why outcome measures are not likely to be good indicators of program impacts. First, outcome measures substantially overstate true impacts since many program registrants would have found jobs or left welfare on their own. Second, the overstatement was not uniform across subgroups. Moreover, outcome differences among programs were determined more by characteristics of the enrollees, local AFDC eligibility regulations and local labor market conditions than by the size of program impacts.

- **The use of longer-term follow-up information about employment and welfare receipt did not improve the correlation of either the job-entry or off-welfare measures with impacts in most cases.**

One possible strategy for improving performance indicators is to increase the length of follow-up data over which the indicators are calculated. Tests of indicators based on longer-term data (up to three years after random assignment) were not successful in this respect. Moreover, the "short-term" outcome measures that were tested made use of follow-up which was already longer than that available to most program operators, who often have only the enrollee's status at date of termination from the program.

- **Weighting outcome measures by prior work and welfare history improved the relationship between performance indicators and impacts in some cases. Still, the development of weighting schemes valid for a variety of program models and local conditions should go beyond this preliminary research.**

Giving more weight to job entries and movement off welfare for individuals with weaker recent work experience improved the correlation between outcome measures and impacts on earnings and welfare receipt. Weighting for longer prior welfare experience also yielded some improvement. Weighting -- whether with simple weights or complex regression-adjustment formulas -- tends to correct the adverse allocation properties of outcome indicators by increasing the incentives to work with less job-ready, more dependent eligibles. At the same time it retains the focus on employment and welfare receipt.

These results should be considered only as one test of the general principal of weighting. The particular weighting schemes tested should not be viewed as the best of all possible schemes. Much remains to be learned about client behavior before a definitive set of variables and weights can be confidently accepted. Moreover, the variety of program approaches and local welfare and labor market conditions may mean that different performance measurement systems will be better in some circumstances than in others.

- **Simple participation measures can also give misleading signals to program operators. Monitoring participation separately for the major subgroups or adopting weighted participation measures may prove more efficient.**

Performance measures based on participation -- that is, activity in program services -- are sometimes proposed as an alternative to job-entry and off-welfare outcome measures. In recent years, participation standards have been criticized as being less directly related to the program goals of employment and welfare departure. Nevertheless, participation measures do have some administrative advantages as indicators of performance. Participation can be readily observed and immediately reported, assisting management in monitoring day-to-day operations. For mandatory programs, monitoring participation may be deemed useful in ensuring consistent treatment of clients across subgroups and local offices. For a program with a potentially large base of eligibles, monitoring participation separately by subgroup can at least ensure that groups with documented impacts are being reached.

Findings of this study suggest that a distinction should be drawn between "maximum participation" and "efficient participation." The subgroup results imply that efforts to maximize total participation may be less efficient than efforts directed towards increasing participation among the moderately dependent (i.e., second tier) subgroups. Extending participation downward into second-tier and third-tier subgroups should increase program effectiveness more than extending participation upward into the first tier. Monitoring participation separately for the major subgroups or weighting participation in the same fashion as outcomes should help achieve that goal.

Conclusions and Unresolved Issues

The research reported here addresses a number of important issues in the monitoring and targeting of welfare employment programs. It also raises questions relevant to the broader employment and training delivery system. The results are suggestive rather than definitive. In some cases, the implications are quite clear. But in others, they raise questions to which the appropriate policy response is less certain.

Subgroup impacts. The findings are clear that, if resources are limited and maximizing program impacts is the goal, it is a mistake to concentrate only on serving the most job-ready portion of the AFDC caseload. Since this was the tendency in the WIN program, this message is an important one and warrants a shift in strategy. The evidence favors the establishment of program goals that encourage working with more dependent and less job-ready individuals. Many administrators are already recognizing this lesson and are adjusting service priorities accordingly.

It is also relatively clear that programs should not focus exclusively on the most disadvantaged among the WIN-mandatory caseload, at least not with low-to-moderate cost services. The results do not provide conclusive guidance to program operators if resource constraints require them to choose among groups of eligibles. There is some evidence that the selection of a targeting approach might depend on the importance attached to different potential program objectives.

Operators who wish to maximize impacts on earnings may find it desirable to work first with applicant returnees or applicants with weak work records. With additional resources, they might next expand services to include longer-term recipients. Operators who seek to maximize welfare savings may want to devote increased effort to the most dependent groups.

The nature of the state initiatives tested, however, suggests a number of cautions in using the results of this study to reach a conclusion on narrow targeting. First, the results reported in this analysis were for programs that did not target narrowly but rather served individuals with a wide range of prior work experience and other factors affecting employability. It is possible -- particularly in group job search components -- that the presence of at least some job-ready enrollees encouraged both program staff and the more disadvantaged, thereby contributing to the positive results reported here.

This "mainstreaming" hypothesis is not tested in this study, but it suggests that administrators should look carefully at the operational results of more targeted services before using resources exclusively for individuals in the second or third tiers. Working only with individuals with lower skills and measured outcomes could have political, administrative or stigmatizing effects. For example, it may be difficult to convince people that a placement rate of 30 percent could represent a substantial positive achievement for more disadvantaged groups, even though this is suggested by the impact findings. Such low rates may also discourage staff efforts. Finally, employers may think differently about a program that refers only clients with no prior work history.

A second reason for caution is that, in allocating resources, program operators have a choice not only in who is served but also in what services are offered. The results reported here suggest that there may be a threshold effect on the earnings (although not the welfare) impacts of low-to-moderate cost programs: They may have comparatively limited impact on the earnings of the most dependent groups. In the future, it will be important to examine whether more intensive services can lead to larger earnings impacts for these individuals. Results from the National Supported Work Demonstration, for example, have shown that earnings increases can be obtained with more extensive services for certain groups with long welfare histories. But this was a small-scale, voluntary program. Its findings may not generalize to more broadly-based interventions.

Performance indicators. This study supports the increasing recognition that alternatives to unadjusted outcome measures are needed to establish valid performance standards. Weighted measures appear to create more appropriate incentives for program operators by explicitly taking account of participants' individual differences. Preliminary evidence suggests that selecting the appropriate characteristics for weighting is at least as important as precision in calculating the weights assigned to those characteristics. When applied to adult welfare recipients, weighting schemes should, at a minimum, include prior employment or welfare history or both.

Nevertheless, even weighted performance indicators have limitations. First, although weighting by demographic charactertisics may correct some of the adverse properties of unweighted outcome indicators, it is not likely to provide a perfect solution to the problem of targeting. For any given enrollee, demographics alone cannot predict with precision who will succeed by participating, and only by participating. The rough guidance provided by weighted indicators may be sufficient for allocating low-cost services; it may not be sufficient for allocating high-cost services. Better data on literacy, general and specific work skills and family circumstances might help identify those who can benefit from particular services, but the potential usefulness and cost of these data are not addressed in this report.

Second, the results presented here come from broad coverage programs charged with enrolling and working with everyone within a specified group of welfare recipients. Very different issues and lessons could arise in selective programs not intended to reach all persons who are categorically eligible. These include programs where participation is voluntary on the part of the client or where local managers can choose the individuals they wish to enroll. In such cases, merely weighting performance measures may have limited effect: Program operators could, for example, screen intensely among the more disadvantaged, identifying only the most able and motivated within the heavily weighted groups, thereby undercutting the objective of serving those who are really less likely to succeed without help.

Third, it will be difficult to develop performance measures that allow meaningful comparisons of effectiveness between programs and across time. A range of factors, such as the local economy and AFDC benefit levels, can have a substantial effect on the composition of the caseload and job prospects for program participants. Comparing program effectiveness with unadjusted measures can be very misleading; further empirical work is needed to determine how reliable such comparisons can be when adjustments are made.

Finally, performance measures are only useful if they can be implemented. The necessary data must be obtainable quickly and at reasonable cost. The calculations must be straightforward enough to be accessible and useful to line staff. The analysis of welfare, and especially earnings, in this report drew on data bases that may only be available to program administrators with considerable lag, if at all. Cost of data collection and ease of interpretation are likely to be important factors in designing measures that are both feasible and valid indicators of true program performance.

CONTENTS

PAGE

ACKNOWLEDGMENTS iii

PREFACE v

EXECUTIVE SUMMARY vii

LIST OF TABLES xxxii

LIST OF FIGURES xxxiv

CHAPTER

1 INTRODUCTION 1

2 THE PROGRAMS AND PARTICIPANTS INCLUDED IN THE STUDY, AND THE NORMAL DYNAMICS OF EARNINGS AND WELFARE RECEIPT 14

3 MEASURING SUBGROUP IMPACT AND ASSESSING PERFORMANCE INDICATORS 44

4 SUBGROUP DIFFERENCES IN IMPACTS 53

5 MEASURE OF PROGRAM PERFORMANCES 98

APPENDIX

A STATISTICAL CONSIDERATIONS IN SUBGROUP COMPARISONS 120

FOOTNOTES 130

REFERENCES 142

LIST OF MDRC STUDIES IN THE WORK/WELFARE DEMONSTRATION 145

LIST OF TABLES

TABLE PAGE

1 SUMMARY OF IMPACTS ON QUARTERLY EARNINGS FOR MAJOR SUBGROUPS OF AFDC APPLICANTS AND RECIPIENTS xiv

2 SUMMARY OF IMPACTS ON QUARTERLY AFDC PAYMENTS FOR MAJOR SUBGROUPS OF AFDC APPLICANTS AND RECIPIENTS xv

2.1 KEY CHARACTERISTICS OF STATE WORK/WELFARE INITIATIVES 17

2.2 DEMONSTRATION OF STATE WORK/WELFARE INITIATIVES: PARTICIPATION AMONG AFDC EXPERIMENTALS 19

2.3 SELECTED CHARACTERISTICS OF AFDC APPLICANTS AND RECIPIENTS AT TIME OF RANDOM ASSIGNMENT, BY PROGRAM AND WELFARE STATUS 25

2.4 AFDC CONTROLS: EMPLOYMENT AND WELFARE RECEIPT BY MAJOR SUBGROUP 34

2.5 AFDC CONTROLS: EFFECTS OF SUBGROUP CHARACTERISTICS ON EMPLOYMENT AND WELFARE RECEIPT 36

2.6 AFDC CONTROLS: NORMAL EMPLOYMENT AND WELFARE BEHAVIOR FOR COMBINATIONS OF THE MAJOR SUBGROUPS 39

2.7 SAN DIEGO AND BALTIMORE: PROGRAM COSTS PER EXPERIMENTAL, BY PROGRAM AND MAJOR SUBGROUP 42

4.1 AFDC APPLICANTS AND RECIPIENTS: UNCONDITIONAL IMPACTS ON EMPLOYMENT, EARNINGS, AFDC INCIDENCE AND PAYMENTS, BY PROGRAM AND WELFARE STATUS 58

4.2 AFDC APPLICANTS AND RECIPIENTS: UNCONDITIONAL IMPACTS ON EMPLOYMENT, EARNINGS, AFDC INCIDENCE AND PAYMENTS, BY PROGRAM, WELFARE STATUS, AND PRIOR YEAR EARNINGS SUBGROUP 61

4.3 AFDC APPLICANTS AND RECIPIENTS: UNCONDITIONAL IMPACTS ON EMPLOYMENT, EARNINGS, AFDC INCIDENCE AND PAYMENTS BY PROGRAM, WELFARE STATUS, AND AFDC HISTORY SUBGROUPS 64

4.4 AFDC APPLICANTS AND RECIPIENTS: IMPACTS FOR SUBGROUPS COMBINING PRIOR EARNINGS, PRIOR AFDC RECEIPT, AND HIGH SCHOOL DIPLOMA STATUS 67

4.5 AFDC APPLICANTS AND RECIPIENTS: UNCONDITIONAL IMPACTS ON EARNINGS AND AFDC PAYMENTS, BY PROGRAM, MINOR SUBGROUP, AND WELFARE STATUS 70

4.6 AFDC APPLICANTS HAVING EARNINGS OF $3,000 OR MORE IN THE YEAR PRIOR TO RANDOM ASSIGNMENT: EMPLOYMENT STATUS IN QUARTERS 4 -- LAST BY WELFARE STATUS IN QUARTER II 81

4.7 AFDC APPLICANTS AND RECIPIENTS: UNCONDITIONAL IMPACTS ON EMPLOYMENT, EARNINGS, AFDC INCIDENCE AND PAYMENTS, BY WELFARE STATUS AND CHILD LESS THAN SIX SUBGROUP 87

4.8 AFDC APPLICANTS AND RECIPIENTS: UNCONDITIONAL IMPACTS ON EMPLOYMENT, EARNINGS, AFDC INCIDENCE AND PAYMENTS, BY WELFARE STATUS, PRIOR EARNINGS, AND PRIOR WELFARE PAYMENTS 91

4.9 AFDC APPLICANTS AND RECIPIENTS: SUMMARY OF IMPACTS FOR MAJOR SUBGROUPS AND COMBINATIONS 92

5.1 AFDC EXPERIMENTALS: VALIDITY OF SIMPLE JOB ENTRY AND OFF-WELFARE PERFORMANCE INDICATORS 102

5.2 AFDC EXPERIMENTALS: VALIDITY AND IMPROVEMENT OF JOB ENTRY AND OFF-WELFARE PERFORMANCE INDICATORS WEIGHTED BY PRIOR EARNINGS 108

5.3 AFDC EXPERIMENTALS: VALIDITY AND IMPROVEMENT OF JOB ENTRY AND OFF-WELFARE PERFORMANCE INDICATORS WEIGHTED BY PRIOR WELFARE 109

A.1 AFDC APPLICANTS AND RECIPIENTS: TESTS OF DIFFERENCES IN IMPACTS ON EARNINGS AND AFDC PAYMENTS BETWEEN SELECTED MAJOR SUBGROUPS 125

LIST OF FIGURES

FIGURE PAGE

2.1 BALTIMORE CONTROLS, EARLY COHORT: QUARTERLY AVERAGE EARNINGS, BY SUBGROUP 30

2.2 BALTIMORE CONTROLS, EARLY COHORT: QUARTERLY AVERAGE AFDC PAYMENTS, BY SUBGROUP 31

4.1 VIRGINIA APPLICANTS: EMPLOYMENT IMPACTS BY PRIOR EARNINGS SUBGROUP 80

SUBGROUP IMPACTS AND PERFORMANCE INDICATORS FOR SELECTED WELFARE EMPLOYMENT PROGRAMS

CHAPTER 1

INTRODUCTION

The search for reliable and workable standards of performance to be used in employment programs for welfare recipients has been one of the major themes in current efforts to reform welfare policy. Such close attention is warranted because performance standards are one of the primary means by which broad policy is translated into the specific objectives that guide the operations of programs. Performance measures fulfill a monitoring function, allowing administrators to assess how well existing programs are doing and to identify problems in the delivery of services. They also fulfill an allocation or targeting function. By encouraging a focus on the welfare groups most likely to help the programs achieve a high performance rating, they influence programs' service priorities and thereby the allocation of funds. To the extent that program objectives include maximizing impacts on individuals, the study of performance measures is bound up with the study of program impacts on subgroups.

This report examines questions about subgroup effects and performance indicators by studying five employment and training programs for recipients of Aid to Families with Dependent Children (AFDC). In all these programs, participation was mandatory. This is the final report of a two-part investigation into the differences among the impacts of these programs on the employment and welfare receipt of selected AFDC subgroups. A previous report presented complete subgroup analyses for only two of the programs, plus a preliminary analysis for the third program.[1]

The study uses data from the Demonstration of State Work/Welfare Initiatives, a five-year, eight-state series of large-scale social experiments conducted by the Manpower Demonstration Research Corporation (MDRC). The data are unusual in that the research samples they describe were generated in controlled experiments based on random assignment. The research focuses on program performance defined in terms of effects on employment and welfare receipt. In addition, an analysis of program costs was undertaken for the principal subgroups in two of the programs.[2] This is reported in Chapter 2.

It should be emphasized that the subgroup study focuses only on AFDC single parents (mostly women) meeting the Work Incentive (WIN) program definition of mandatory: These were parents who had no child under the age of six (under three in Arkansas) and no other known barriers to participation. This WIN-mandatory group makes up just over one-third of the AFDC caseload nation wide. Unemployed heads of two-parent households who are also WIN-mandatory were typically part of the MDRC research samples in the states that had an AFDC-U program and therefore served this group. These samples have been excluded from the subgroup study, however, because their behavior is typically too different from the AFDC group to be analyzable as part of that group. They receive assistance under different rules from those applying to the AFDC single-parent case heads, for example; they have different labor supply patterns; and, because WIN-mandatories under AFDC-U are primarily married men, they have different work backgrounds. The Arkansas program extended mandatory status to AFDC case heads with a youngest child aged three to five. The results for this group are included, and are of particular interest in the context of the current policy debate.

The implications of the analysis are somewhat broader in scope than mandatory welfare employment programs because many of the issues examined are common to other programs for low-income or disadvantaged groups, such as those funded by the Job Training Partnership Act (JTPA). But some care should be taken in generalizing from the conclusions. In particular, it should be borne in mind that the programs studied were usually intended for mass coverage, and the cost per enrollee was relatively low. Performance standards and targeting strategies may differ for services that have higher per enrollee costs.

In some respects, the role of targeting has less to do with the voluntary/mandatory distinction per se than with the extent to which the program is given discretion to select its participants. Programs expected to serve all or substantially all those who are technically eligible will have less opportunity for targeting than programs whose primary purpose is to maximize impacts on earnings or welfare receipt for only a portion of the eligible population, for example if resources are not sufficient to serve everyone. In the former case, the "requirement" nature of the program applies not only to the clients but also to the operators if they are required by law to enroll all eligible persons.

For such broad coverage programs, selective targeting may be inconsistent with program objectives. The goal would not be to narrow the focus of a program, but rather to extend it to groups of eligibles who have had low participation rates in the past. For example, a mandatory "workfare" program, whose goal is the payback of welfare benefits through community service by all able-bodied recipients, may monitor particular groups to ensure that inequities are not created through unequal application of the participation requirement. The program may also target a group for special attention to raise its participation rate if it is below average.

In contrast, selective coverage programs working with a narrow subgroup of welfare recipients expected to reap the largest benefits from the program may be mandatory, but are typically voluntary programs. Because selective or partial coverage programs are not intended to serve everyone, targeting may be the paramount design question, and performance standards that carry implicit incentives for targeting in a particular fashion may be one of the primary influences on how a program is implemented and what its impacts on participants are. In addition, if resources are too limited to serve everyone, targeting may help programs to use resources efficiently.

This report first examines differences in impacts across subgroups that vary in their prior employment and welfare history, and in a variety of individual characteristics (education, age, number of children, and the like). The intent of the comparison is to identify whether the relative impacts on any subgroups are consistent across the five programs. To the extent that they are, more general conclusions can be drawn about which subgroups benefit the most from the types of programs included in the evaluation.

The analysis then uses the subgroup impacts generated from the experimental data to evaluate the validity of two frequently used performance measures -- the number of "job entries" (placements)[3] and the number of cases "off-welfare" (case closures) -- and variants of them. The issues addressed in this report also apply to other measures of performance, such as wage rates, job retention, and program participation.

The discussion is structured as follows. The rest of this chapter reviews issues relevant to welfare population subgroups and program performance. Chapter 2 discusses the welfare employment programs studied and their research designs. Chapter 3 explains the methodology used to estimate subgroup impacts and to test performance indicators. Chapter 4 presents impacts for the major subgroups in the study. Chapter 5 evaluates the validity of several alternative formulations of employment- and welfare-related performance measures, using program impact estimates.

I. Issues in Program Performance

Recent research on relatively low-cost welfare employment programs has put issues regarding program performance into sharp focus. Earlier studies as part of the MDRC Work/Welfare Demonstration are illustrative. Local experiences vary widely, but typical employment rates for a group of prospective enrollees without the program might average about 30 percent over a three-month period. The programs cost in the range of $150 to $1,000 per experimental. After the programs were established, about half of those enrolled typically participated in some formal activity. The overall employment effect on the enrollee groups, counting nonparticipants as well as participants, averaged an increase in the employment rate of around 4 percentage points, from a 30 to a 34 percent employment rate.

Not all who are enrolled in a program benefit from it. Some who find jobs would have found them without the program. Conversely, some who participate do not benefit. Not all who are given the chance to participate do so. Some leave welfare for reasons unrelated to program participation or even to employment (such as getting married or moving). One of the possible goals of targeting, therefore, is to concentrate resources on the groups most likely to benefit from the program, or most cost effective to serve.

A. Targeting

Much of the recent work in targeting welfare employment programs has focused on AFDC subgroups outside the WIN-mandatory category -- such as mothers with young children.[4] It has been found that length of welfare dependency is related to objectively measured individual characteristics and that the majority of people who enter the welfare system spend less than four years on the rolls, even counting repeat spells. The minority of people who remain on welfare for many years account for the bulk of AFDC benefit expenditures; one study estimated that as much as 60 percent of all grant outlays are paid to only 25 percent of all recipients.[5]

Young mothers, particularly never-married mothers, appear to be at especially high risk of long-term dependency. The conditions and problems that lead to extended dependency, however, may not be amenable to change with low-cost employability services.[6] In any case, the study has only limited ability to address this most dependent group and the services that may be most effective for them for two reasons. First, they are typically not in the traditional WIN-mandatory category, the focus of the programs being compared here. "Most-dependent" as used in this report refers only to the most dependent of those among the WIN-mandatory portion of the AFDC caseload. Only one of the programs studied here includes substantial numbers of women with children under six years old. Second, the programs did not emphasize some services (such as day care) that might be of particular value to working women with preschool children.

The results of this study are most useful for planning programs similar to those evaluated here: relatively low-cost services designed for the WIN-mandatory portion of the AFDC caseload. In program planning, the cost per person for a particular kind of intervention will normally be a given. If this is combined with a fixed total budget, the number of persons who can be served is also determined by the following simple identity:

NUMBER OF PERSONS SERVED = (TOTAL COST)/(COST PER PERSON).

In cases where the number that can be served is smaller than the total number eligible, information from a study such as this about relative program effectiveness across different subgroups may be useful in deciding who should be given priority for services.

It should be noted that narrowing the target group within the WIN-mandatory population carries many uncertainties. Resources may be freed up, but restricting the target group may reduce the total effect of services. This is because the impact of a program will automatically be zero for those who are excluded from it. The total effect of a service is given by the simple identity:

TOTAL IMPACT = (IMPACT PER PERSON) X (NUMBER OF PERSONS).

Targeting may very well increase the impact per person but, by definition, decreases the number of persons served. Targeting one-half of the eligibles, for example, means that average impact per target group member must be twice the average impact for all persons taken together to achieve the same overall impact.

There is even some chance that narrow targeting -- e.g., on the more disadvantaged -- may reduce program effects on the target groups. This is equivalent to the question of "tracking" versus "mainstreaming," so prominent in education. An open question in welfare employment programs, for example, is whether loosely structured, low-cost services, such as job search workshops, can be as effective if women with no prior work experience do not have the opportunity to learn from others who have held jobs.

B. Measures of Performance

Performance measures are intended to promote program effectiveness, conserve resources, and ensure compliance with overall goals and directives. Both in monitoring and in the allocation of funds, performance standards play a critical role in determining the efficiency of program expenditures; and, in a period of fiscal restraint, it is particularly important to choose performance measures that increase rather than decrease efficiency. Poorly designed or inadequately tested performance standards can work against the objectives of the authorizing legislation. They can promote methods of operations that waste staff time and other program resources, with the result that neither the welfare population nor society is well served.

A wide range of indicators has been developed and used in the WIN program, programs funded by the Comprehensive Employment and Training Act (CETA) and, more recently, programs funded by JTPA. Historically, job placements and welfare reductions have been the most important indicators in WIN. These measures have seemed useful in conveying program achievements in straightforward terms to policymakers and the general public. Their incorporation into the fiscal WIN Allocation Formula underlined their significance to operators of welfare employment programs. Other indicators were also part of the WIN Allocation Formula, such as the quality of job entries, usually measured by wage rates and job retention.[7]

These indicators all measure the outcomes of a registrant's program experience at some point after registration. Another set of indicators looks at the activity of registrants while in the program; these include counts of registrants, participants, program completers and similar measures. Participation data have been examined in evaluations of WIN, CETA and other programs. The trend recently has been to deemphasize these indicators, even though they provide immediate feedback and the required information is relatively inexpensive to collect.[8] Instead, emphasis has been on measures that communicate program goals in terms of post-program outcomes. For example, the JTPA legislation explicitly requires that standards for adult participants be based on job entries, wages and earnings, retention and welfare reductions.[9]

C. Outcomes and Impacts

The distinction between outcomes and impacts is critical to an understanding of program performance. An outcome is the employment and/or welfare status of a person at some point in time after program registration. Hence, the outcome "employed and not receiving welfare at quarter 4" describes the status of a person 9 to 12 months after program entry. The real effects of a program cannot be judged by outcomes, however, given the high degree of normal job-finding and welfare departure within the welfare

44

population (i.e., outcomes that are not related to program experience).

Program impact is the effect of the program itself. It is the difference between outcomes with the program and outcomes without it. This study estimates program impacts as the difference between the outcomes of a randomly selected group of people eligible for the program treatment (the experimental group) and the outcomes of a similar group of people not eligible for the treatment (the control group). The distinction between a level (the outcome) and a difference (the impact) is fundamental.

Past research has suggested that groups exhibiting worse-than-average outcomes may, in fact, experience better-than-average program impacts. For example, an MDRC evaluation of a job search and work experience program operated in San Diego found that 73 percent of WIN-mandatory AFDC applicants who had worked at some time during the year prior to their program entry were able to find employment during the 18 months following enrollment. This outcome was, in fact, only a 2 percentage point increase (or impact) over the control group employment outcome -- that is, the rate that applicants with a prior work history were able to achieve on their own. In contrast, of the enrollees without prior employment only 48 percent were able to find employment; but this outcome was a 10 percentage point increase over the control group's employment rate of 38 percent.[10]

Given these patterns, performance indicators based only on outcomes create a misleading impression of program effectiveness. They may also lead to ineffective targeting of program resources if these standards place emphasis on serving the least appropriate groups -- that is, those who would have done well on their own without the program. Conversely, people who could benefit most from these programs may be underserved.

D. Evaluating Performance Indicators in Light of Impact Differences

Historically, all enrollees have been given equal weight in rating performance, whatever the measure used. The findings in this report and similar ones from other studies, however, suggest that consideration be given to the development of performance formulae that allow outcome standards to vary by local economic conditions, registrant characteristics, and even by service components. Regression adjustment to control for these variations, as is done in the JTPA formulae, has the advantage of permitting more flexible performance standards for programs serving groups with a low likelihood of finding employment readily or high service costs, or programs operating in labor markets where it is hard to find jobs.

But regression formulae also have disadvantages. The formulae can be complex, making them unsuitable for communication of program objectives to local staff or for setting performance criteria for service subcontractors. Moreover, the regression procedure must be properly executed: a formula must include the most important determinants of outcomes and must be estimated from an appropriate sample if its message is to be correct.

This study presents some simpler formulae, which weight individual performance indicators by prior employment and prior welfare receipt (on the basis of the subgroup impact analysis), the characteristics used to define the major subgroups. These measures were chosen because they are both important predictors of future behavior for adult WIN-mandatory AFDC women, as this report will show. This choice does not imply that other approaches have been rejected. The study recognizes that no single approach should necessarily be applied to all programs and labor markets. It recognizes that the goals of some programs may not be easily translated into any single formula, whether simple or complex. Moreover, different goals may require different formulae.

II. The Demonstration of State Work/Welfare Initiatives

MDRC's Demonstration of State Work/Welfare Initiatives was launched in 1982 to test the effectiveness of state employment programs for people applying for or receiving AFDC. For the most part, states were using their new authority to experiment with WIN program variations authorized by the Omnibus Budget Reconciliation Act (OBRA) of 1981. The MDRC study includes programs in 11 states, eight of which used random assignment to form experimental and control groups for full-scale impact and benefit-cost studies. Most programs set the goals of increasing employment and reducing the dependency of the welfare population by preparing recipients for work. They required most able-bodied recipients to participate in job search and/or unpaid work experience or other activities as a condition of welfare receipt.

The research was designed to assess three areas: the feasibility of implementing a mandatory participation and/or work requirement; the programs' impacts on employment, earnings and welfare receipt; and the cost-effectiveness of the different approaches. Findings from this MDRC demonstration are being released as the results for each state's program become available. The five initiatives included in this subgroup study are described in Chapter 2.

Among the five program evaluations on which this report is based, four found positive employment and earnings impacts. Four of the five also obtained short-run welfare reductions. And four of the five indicated that the initial investment of program funds would result in net government budget savings in five years or less.

The individual evaluations included some research on subgroup impacts. Those findings suggest that programs may indeed have greater impacts for some groups within the diverse welfare population than for others. For example, employment increases have often been larger for clients without a recent work history than for those who have worked during the year prior to program enrollment. These findings are buttressed by MDRC research on prior WIN programs and the results from the National Supported Work Demonstration.[11] The study reported on here is able to examine a wider variety of subgroups than were analyzed in the individual evaluations. It also uses longer-term data and a methodology more suited to the issue of performance measures.

CHAPTER 2

THE PROGRAMS AND PARTICIPANTS INCLUDED IN THE STUDY, AND THE NORMAL DYNAMICS OF EARNINGS AND WELFARE RECEIPT

To provide some context for the discussion of program impacts in Chapter 4, this chapter describes the programs included in the evaluation, the characteristics of the registrants who were offered the opportunity to participate, and the welfare and employment experiences they would have had in the absence of program services. The first section discusses key similarities and differences among the programs included in the subgroup study: San Diego, Baltimore, Virginia, Arkansas, and Cook County, Illinois (which includes the City of Chicago). The second section describes the characteristics of sample members. The third section looks at the earnings and welfare receipt patterns for different subgroups of controls. The fourth section discusses how the major subgroups are defined for the analysis. The final section illustrates the differences in program costs for different subgroups by comparing costs of the San Diego and Baltimore programs.

I. The Program Models

No single program model was tested in MDRC's Work/Welfare study. Rather, the participating states implemented their own initiatives, using different strategies. Characteristics of the local WIN-mandatory populations often differed as well.

The evaluations, on the other hand, are similar in methodology: Each study used an experimental design whereby program eligibles were randomly assigned to one or more experimental groups or to a control group. Members of the experimental groups were required to take part in the program services being evaluated, whereas the control groups were barred from the special program services, although in some areas they could receive the minimal services offered under the regular WIN program. Data were collected on participation measures, employment and welfare outcomes, and direct program operating costs. To estimate program impacts, the employment and welfare behavior of the experimental and control groups were compared over several quarters of follow-up. Because random assignment produced experimental and control groups with similar demographic characteristics and backgrounds in prior employment and welfare dependency, any statistically significant differences in behavior during the follow-up period can be confidently attributed to the program.

The term applicant identifies a person applying for AFDC at the time of entry into the research sample, whether or not that person's welfare grant was subsequently approved. That label remains, even for those applicants who never get approved, and even for those applicants who become recipients. The term recipient refers to a sample member who was already receiving welfare at the date of sample entry. These two subgroups are important and are analyzed separately throughout this study. Other subgroup divisions are based on prior demographic and background characteristics.

Table 2.1 shows the key characteristics of the programs included in this analysis. Table 2.2 shows rates of participation for experimentals in the various components. Length of follow-up is important for interpreting participation rates. In these samples the follow-up for participation is different from the follow-up for the impact data and is shown for each program in the bottom row of the table. The published state reports contain more detail about both the programs and the evaluation results.[1,2]

Briefly, job search and work experience were the major program services; but states differed in the mix and intensity of these services, their sequencing, and the populations that received them. Some participation in education and training was recorded for Baltimore, Virginia, and Cook County. But in the latter two this participation largely reflected referrals to outside providers or self-initiated activity and was little or no higher than the background levels observed for controls. The proportion participating in any program activity varied from 38 percent in Arkansas to 58 percent in Virginia. The programs were all mandatory, but differed in the extent to which participation was enforced and the degree to which monetary sanctions were used as a tool of enforcement. The proportion sanctioned varied from practically zero in Baltimore to 12 percent in Cook County.

San Diego worked with all WIN-mandatory welfare applicants but did not enroll recipients. Experimentals went through a two-stage fixed sequence of group job search, followed by a 13-week work obligation if they had not found unsubsidized jobs in the first phase.[3] San Diego's decision to focus entirely on applicants represents one targeting option available to program operators.

Baltimore enrolled both WIN-mandatory applicants and recipients, but only recipients who had just become mandatory, usually because their youngest child had turned six years of age. The program provided a mix of components (including job search, unpaid work experience, education and

TABLE 2.1

KEY CHARACTERISTICS OF STATE WORK/WELFARE INITIATIVES

Characteristic	San Diego, California[a]	Baltimore, Maryland[b]
Eligible Group		
Applicants	Yes	Yes
Newly Mandatory Recipients	No	Yes
Currently Mandatory Recipients	No	No
Enrollment Limit	None	1000/year
Program Model	Job search workshop followed by 13 weeks of CWEP in public and private nonprofit agencies.	Multi-component, including job search, education, training, on-the-job training and 13 weeks of work experience.
Sequence	Fixed: job search then work experience	Discretionary
Client Choice of Components	No	Yes
Components		
Job Search	Mandatory	Mandatory when judged appropriate
Independent	No	Yes
Group	Yes	Yes
Work Experience	Mandatory if no job found through job search	Mandatory when judged appropriate
Education and Training	Not offered	In-house and by referral
Study Area[c]	County-wide	10 out of the 18 Income Maintenance Offices
Control Services	WIN services	WIN services
Sample Enrollment Period	October 1982 - August 1983	November 1982 - December 1983

TABLE 2.1 (continued)

Virginia[a]	Arkansas[b]	Cook County, Illinois[a]
Yes	Yes	Yes[d]
Yes	Yes	Yes
Yes	Yes[e]	Yes[e]
None	None	None
Job search followed by 13 weeks of CWEP, education or training.	Job search workshop followed by individual job search and 12 weeks of work experience in public and private non-profit agencies.	Individual job search followed by 13 weeks of CWEP in private non-profit agencies.
Job search first	Job search first	Job search first
Yes, after job search	No	No
Mandatory as first component	Mandatory	Mandatory
Yes	Yes	Yes
Yes	Yes	No
Mandatory when judged appropriate	Mandatory, but used infrequently	Mandatory if no job found through job search
By referral	Not offered	Not offered
11 of 124 agencies (4 urban, 7 rural)	Pulaski South and Jefferson Counties	16 out of the 22 Income Maintenance Offices
No special services	No special services	Attend the orientation session
August 1983 - September 1984	June 1983 - March 1984	February 1985 - November 1985

NOTES: [a] In San Diego, Virginia, and Cook County, there are two different experimental treatments. In Virginia, the two experimental groups were merged for the analysis.

[b] In Maryland and Arkansas, a full evaluation was conducted in the indicated counties and a process study covered other areas as well.

[c] In addition to the study areas, Virginia and Illinois implemented their programs statewide and Arkansas and Maryland in selected areas.

[d] Unlike other states, applicants in Cook County were all approved for AFDC before enrollment.

[e] Although "currently mandatory recipients" were eligible for the program, this group was not included in the research sample.

TABLE 2.2

DEMONSTRATION OF STATE WORK/WELFARE INITIATIVES:
PARTICIPATION AMONG AFDC EXPERIMENTALS

Program Activity Measure	San Diego	Baltimore	Virginia[a]	Arkansas	Cook County
Participation Rate					
Any Activity (%)[b]	44.6	45.0	58.3	38.0	47.2
Job Search					
Individual (%)	---	24.7	40.4	23.3	36.1
Group (%)	42.3		14.7	27.3	---
Work Experience (%)	11.8	17.5	9.5	2.9	7.3
Education and Training[c] (%)	4.1	17.3	11.6	2.4	16.9
Deregistered (%)	52.1	37.6	42.3	57.5	56.9
Due to Request for Sanctioning (%)	6.6	rare	3.8	4.3	12.4
Sample Size	1540	1362	2138	245	4050
Follow-Up Period in Months	Six	Twelve	Nine	Nine	Nine

SOURCE: MDRC Work/Welfare Demonstration Reports.

NOTES: [a] For Virginia, activity measures are based on both experimental groups, which differed in intended access to work experience and education and training activities.

[b] For San Diego and Cook County, activity measures refer to post-registration or post-orientation activities.

[c] Includes services other than education and training in San Diego and Arkansas.

training), and staff made service assignments taking into account enrollees' needs and preferences, depending on their assessments and the availability of open slots. In order to ensure adequate funding on an individual basis for this somewhat broader array of services than offered in the other programs, the Baltimore program restricted active enrollment to only 1,000 registrants per year during the period studied.

Virginia enrolled its entire WIN-mandatory caseload. The state stipulated that counties require job search of all enrollees but authorized, as a county option, short-term work experience, education and training as follow-up activities. Education and training were not provided by the program; rather, participants were referred to JTPA and community schools with independent funding, open to all who qualified. As it turned out, enrollees participated in education and training at the same rate as controls.

Arkansas enrolled its entire WIN-mandatory caseload during the program start-up phase, but only applicants and recipients who became mandatory after the research began were included in the impact sample. The program consisted primarily of independent and group job search and, less frequently, a work experience component. Three features distinguish the Arkansas sample. First, the state has relatively low AFDC grant levels. Individuals are therefore likely to apply for AFDC if they have very little opportunity for income through work or through other family members. As a consequence, employment rates -- both before and after enrollment -- were lower for this sample than for the others. Second, Arkansas had the largest share of applicants whose grants were not approved. For this reason alone welfare receipt will be lower for this sample than for other samples included in the study.

Third, Arkansas, under federal waiver provisions, filed for and received permission to classify as WIN-mandatory AFDC case heads whose youngest child was aged three to five years. As stated in Chapter 1, much of the descriptive analysis of potential target groups has focused on women with children younger than school age. The Arkansas sample affords an opportunity to examine actual program effectiveness on an important part of this subgroup, albeit with a small sample and at grant levels below those of most states.

The Cook County program only worked with recipients and a subset of applicants, those whose AFDC grants had already been approved. As with Arkansas, the Cook County research sample was restricted to those who became WIN-mandatory after the research began. Sample members were expected to participate in independent job search for two months. They were required to make 40 employer contacts per month and to report on these at biweekly group sessions. A stock of short-term unpaid work experience positions was also maintained, and individuals who completed job search without a job could be assigned to a worksite when one opened.[4] In addition to these program activities, experimentals were allowed to participate in education and training activities they might find on their own. Finally, compared to other programs, staff were more oriented towards obtaining welfare reductions from reported client employment and sanctioning enrollees who failed to carry through on their assignments than toward service provision.

These five programs were all relatively inexpensive, although they varied somewhat in average cost. For example, the net cost of the San

Diego program (in 1987 dollars) was about two-thirds that of the Baltimore program -- which spent, on average, $1,050 per experimental. Costs for Virginia (around $450 per experimental) were in the middle of the range for the five programs. Arkansas and Cook County cost the least, at $170 and $150 per experimental, respectively. Programs also differed in where they chose to concentrate resources. San Diego spent more on ensuring compliance with its participation requirement (which entailed monitoring, registrant follow-up and limited sanctioning), for example, whereas Baltimore offered more expensive services, such as education and training, and provided client stipends.

The programs also differed in the proportion of eligibles covered. This is different from the proportion that participated because it also includes those who became employed, left welfare for other reasons, or were sanctioned. People were defined as not covered by the program if they were still on welfare, not employed, and had not participated in a program activity nine months after program entry. In San Diego, for example, the participation rate was 45 percent, but only 10 percent were not covered by the program within nine months of enrollment. This high San Diego coverage indicates that a short-term participation requirement was, in fact, realized by that program. In contrast, a larger proportion of continuing registrants -- almost one quarter -- were not covered in the Baltimore program. This may be due partly to inclusion of new WIN mandatory recipients in Baltimore and the staff's greater flexibility in deferring registrants from activities. In Virginia, most experimentals were covered by the program (nearly 90 percent), but the minimum requirement -- a loosely structured form of independent job search -- was relatively easy for both the program and the clients to fulfill. Arkansas and Cook County, the two programs with lowest costs per enrollee, had different coverage rates. Participation rates were somewhat lower in Arkansas than elsewhere, but approval rates for applicants were also lower, yielding an overall coverage rate that was similar to Baltimore's. Cook County, despite its focus on recipients and despite its large caseload, reached 94 percent coverage by using independent job search as the main component, backed by routine sanctioning for noncompliance.

Statutory grant maximums, based on state standards of need, varied widely, requiring extra care in making comparisons among programs. Low benefit levels increased the attractiveness of low-wage jobs in some areas, and also increased the likelihood of a case closure when employment was obtained. Local economic conditions, staff experience and attitudes also differed. In San Diego, welfare recipients had a relatively good market in which to look for jobs, but in rural areas of Virginia, the prospects for employment were more limited. And, in the administrative reorganization permitted by OBRA, social service staffs in some states -- those who had not had responsibility for employment functions under the previous system -- had to go through a learning process. However, staffs in San Diego and Baltimore had substantial prior experience in operating such programs, which contributed to their programs' smooth administration.

As shown in Table 2.1, services available to controls varied across experiments. In Virginia and Arkansas, no special services were provided. In Cook County, controls were required to attend only the initial orientation meeting, but were not assigned to job search or work experience. In San Diego and Baltimore, controls were assigned to the existing WIN programs, although services there were low. For example, participation rates among Baltimore controls were under 5 percent.

II. Sample Characteristics

The sizes and demographic composition of the research samples for the five programs analyzed in this report are shown in Table 2.3.[5] The evident variation is due to differences in the program models, targeting philosophies, and environments in which the programs operated. As noted earlier, each program served the WIN-mandatory caseload or portions of that caseload. In four of the five programs, this meant excluding most women with children less than six years old. Arkansas extended WIN-mandatory status to AFDC women whose youngest child was aged three to five. The San Diego program served only applicants, while the Baltimore, Virginia, and Arkansas samples had a fairly even mix of applicants and recipients, although the type of recipient differed. Cook County worked with mandatory applicants and recipients, but only applicants whose AFDC grant had been approved.

There were other differences not shown in Table 2.3. AFDC approval rates for Arkansas applicants were lower than elsewhere, for example, partly because program enrollment occurred at the time of the welfare application rather than later. Also, as indicated earlier, employment rates in Arkansas were low compared to the other states, probably reflecting the low grant levels there.

Analysis of the Cook County data proved to be the most difficult to integrate with the other programs. Applicant status was not available in the data set as it was in other states and had to be inferred from the

TABLE 2.3

SELECTED CHARACTERISTICS OF AFDC APPLICANTS AND RECIPIENTS AT TIME OF RANDOM ASSIGNMENT, BY PROGRAM AND WELFARE STATUS

	San Diego	Baltimore			Virginia		
Subgroup	Applicants	Applicants	Recipients	Total	Applicants	Recipients	Total
Research Group (%)							
Experimental	46.5	48.6	50.2	49.4	66.7	67.7	67.3
Control	27.0	51.4	49.8	50.6	33.3	32.3	32.7
Other	26.5	---	---	---	---	---	---
Prior Earnings (%)							
$3000 or More	28.8	31.9	6.7	19.3	29.2	3.3	13.7
$1-2999	22.9	29.3	20.6	24.9	28.4	19.7	23.2
None	48.4	38.8	72.8	55.8	42.3	77.0	63.0
Had Own AFDC Case (%)							
Never	33.4	22.7	5.2	14.0	26.2	2.5	12.0
Two Years or Less	38.7	41.8	21.1	31.4	31.7	25.7	28.1
More Than Two Years	27.9	35.5	73.8	54.6	42.2	71.8	59.8
High School Diploma (%)							
Yes	61.5	44.9	42.1	43.5	50.8	38.8	43.6
No	38.5	55.1	57.9	56.5	49.2	61.2	56.4
Child 12 or Under (%)							
No	22.6	27.5	13.4	20.5	22.9	23.7	23.4
Yes	77.4	72.5	86.6	79.5	77.1	76.3	76.6
Number of Own Children (%)							
One	49.7	50.4	43.1	46.8	49.6	42.0	45.0
More Than One	50.3	49.6	56.9	53.2	50.4	58.0	55.0
Currently Married (%)							
Yes	46.6	50.4	34.3	42.3	49.3	38.3	42.8
No	53.4	49.6	65.7	57.7	50.7	61.7	57.2
Ever Married (%)							
Yes	84.1	69.9	49.1	59.5	74.2	65.3	68.9
No	15.9	30.1	50.9	40.5	25.8	34.7	31.1
Age (%)							
30 or Over	65.6	65.4	42.7	54.0	64.0	65.9	65.1
Under 30	34.4	34.6	57.3	46.0	36.0	34.1	34.9
Ethnicity (%)^{c}							
White	61.5	33.8	25.1	29.5	41.8	26.8	32.8
Black	20.7	66.2	74.9	70.5	58.2	73.2	67.2
Hispanic	17.8	---	---	---	---	---	---
Recent UI Benefits (%)							
Some	14.1	---	---	---	3.9	0.4	1.8
None	85.9	---	---	---	96.1	99.6	98.2
Labor Market (%)							
Urban	---	---	---	---	78.7	78.8	78.8
Rural	---	---	---	---	21.3	21.2	21.2
Sample Size^{d}	3238	1380	1377	2757	1269	1881	3150

(continued)

TABLE 2.3 (continued)

Subgroup	Arkansas			Cook County[a,b]		
	Applicants	Recipients	Total	Applicants	Recipients	Total
Research Group (%)						
Experimental	50.0	49.2	49.7	34.5	33.8	34.0
Control	50.0	50.8	50.3	32.0	31.9	31.9
Other	---	---	---	33.5	34.3	34.1
Prior Earnings (%)						
$3000 or More	12.5	1.3	8.0	33.0	4.9	14.4
$1-2999	22.5	7.2	16.3	20.4	14.9	16.8
None	64.9	91.5	75.7	46.5	80.2	68.8
Had Own AFDC Case (%)						
Never	56.3	8.1	36.7	---	---	---
Two Years or Less	36.7	26.9	32.7	---	---	---
More Than Two Years	7.0	65.0	30.5	---	---	---
High School Diploma (%)						
Yes	55.1	42.0	49.8	46.8	30.9	36.2
No	44.9	58.0	50.2	53.2	69.1	63.8
Child 12 or Under (%)						
No	9.6	11.8	10.5	---	---	---
Yes	90.4	88.2	89.5	---	---	---
Number of Own Children (%)						
One	42.5	34.8	39.4	---	---	---
More Than One	57.5	65.2	60.6	---	---	---
Currently Married (%)						
Yes	31.9	21.2	27.6	---	---	---
No	68.1	78.8	72.4	---	---	---
Ever Married (%)						
Yes	56.9	41.8	50.8	---	---	---
No	43.1	58.2	49.2	---	---	---
Age (%)						
30 or Over	38.5	38.1	38.3	62.1	43.1	49.5
Under 30	61.5	61.9	61.7	37.9	56.9	50.5
Ethnicity (%)[c]						
White	16.7	8.5	13.4	21.8	14.3	16.8
Black	83.3	91.5	86.6	65.5	75.3	72.0
Hispanic	---	---	---	12.6	10.4	11.1
Recent UI Benefits (%)						
Some	---	---	---	---	---	---
None	---	---	---	---	---	---
Labor Market (%)						
Urban	64.8	56.5	61.4	---	---	---
Rural	35.2	43.5	38.6	---	---	---
Sample Size[d]	670	457	1127	4014	7898	11912

SOURCE: Demographic information is from MDRC Client Information Sheets. Prior earnings were calculated from the County of San Diego Unemployment Insurance records from the EPP Information System; from the Commonwealth of Virginia Unemployment Insurance earnings records; from State of Arkansas Unemployment Insurance records; and from the Illinois Unemployment system earnings records.

Table 2.3 (continued)

NOTES: Distributions may not add to exactly 100.0 percent due to rounding. Categories not applicable for particular program samples are indicated with a dash. Tests of statistical significance were not calculated.

[a] Cook County demographic information is from the State of Illinois Department of Public Aid AAID system records.

[b] The definitions of "applicant" and "recipient" for Cook County differ from the other programs. See text for discussion.

[c] For Baltimore, Virginia, and Arkansas the category "black" includes a small number of individuals in other non-white groups. In San Diego and Cook County, "white" includes a small number of non-black, non-hispanic, non-white persons.

[d] There were two experimental groups in San Diego and Cook County. The percent of sample calculations are based on Job Search/Work Experience Experimentals, Job Search Only Experimentals and Controls, for a total of 3238 for San Diego and 11912 for Cook County. Impact calculations do not include Job Search Only Experimentals; the sample sizes for San Diego and Cook County in the impact analysis are, therefore, 2381 and 7855, respectively.

absence of AFDC payments prior to random assignment. In addition, the fact that only individuals whose AFDC applications had already been approved were enrolled in the program meant that "applicants" were more likely to be on welfare during follow-up than applicants in the other samples. More serious, information about the length of prior welfare receipt and other characteristics for Cook County is lacking because the information sheets used to obtain subgroup characteristics did not cover the full impact sample in Cook County.

The Baltimore and Virginia samples were similar in many respects: over half had neither a high school diploma nor a GED; more than half had been receiving AFDC for more than two years; and, on average, only about 40 percent had held a job in the year prior to random assignment. The San Diego sample was less disadvantaged: more than half were high school graduates; less than 30 percent had been on welfare for more than two years; and one-half had held a job in the year before this welfare application.

Ethnic composition also differed. In Baltimore and Virginia, between 60 and 70 percent of the samples were black; in Arkansas, over 85 percent were black; in San Diego only 20 percent of sample members were black, with Hispanics making up 18 percent of the sample. In Cook County, more than two-thirds of the sample members were black, about 11 percent were Hispanic. Data on educational attainment in Cook County, not strictly comparable to the data for the other programs because they came from case records rather than direct interview, indicate that nearly two-thirds of the sample did not have a high school diploma. Arkansas was characterized by particularly low rates of prior-year employment, under 25 percent. In addition, prior welfare histories among Arkansas sample members varied widely, with about a third of the sample never having had an AFDC case in the past and a third having had a case for more than two years.

Comparisons of applicants and recipients reveal large differences in prior earnings and prior welfare receipt in all programs. Applicants not only had shorter welfare histories, but also had more recent earnings and, except for Baltimore, more education. Recipients were more likely to be unmarried at enrollment and also more likely never to have been married. At least two-thirds of recipients in the three samples with data had a welfare history of more than two years.

III. Earnings and Welfare Receipt: The Normal Dynamics

A wide range of earnings and welfare information on WIN-mandatory clients in the absence of special program intervention can be captured by simple objective measures, collected as part of the program enrollment process and readily verifiable. As an illustration, Figures 2.1 and 2.2 plot the earnings and welfare receipt of the early Baltimore control sample over the three-year period after random assignment, for selected subgroups defined by applicant/recipient status, length of prior welfare receipt, and prior earnings. The Baltimore sample was selected because it has the full spectrum of applicant and recipient subgroups and a long enough follow-up period to show the importance of changes in status over time.

The subgroup differences are typically large. Quarterly average earnings for control group applicants without a prior welfare history and with $3,000 or more in earnings in the year prior to AFDC application (i.e., the point of random assignment), for example, fall into the $1,200

FIGURE 2.1
BALTIMORE CONTROLS, EARLY COHORT: QUARTERLY AVERAGE EARNINGS, BY SUBGROUP

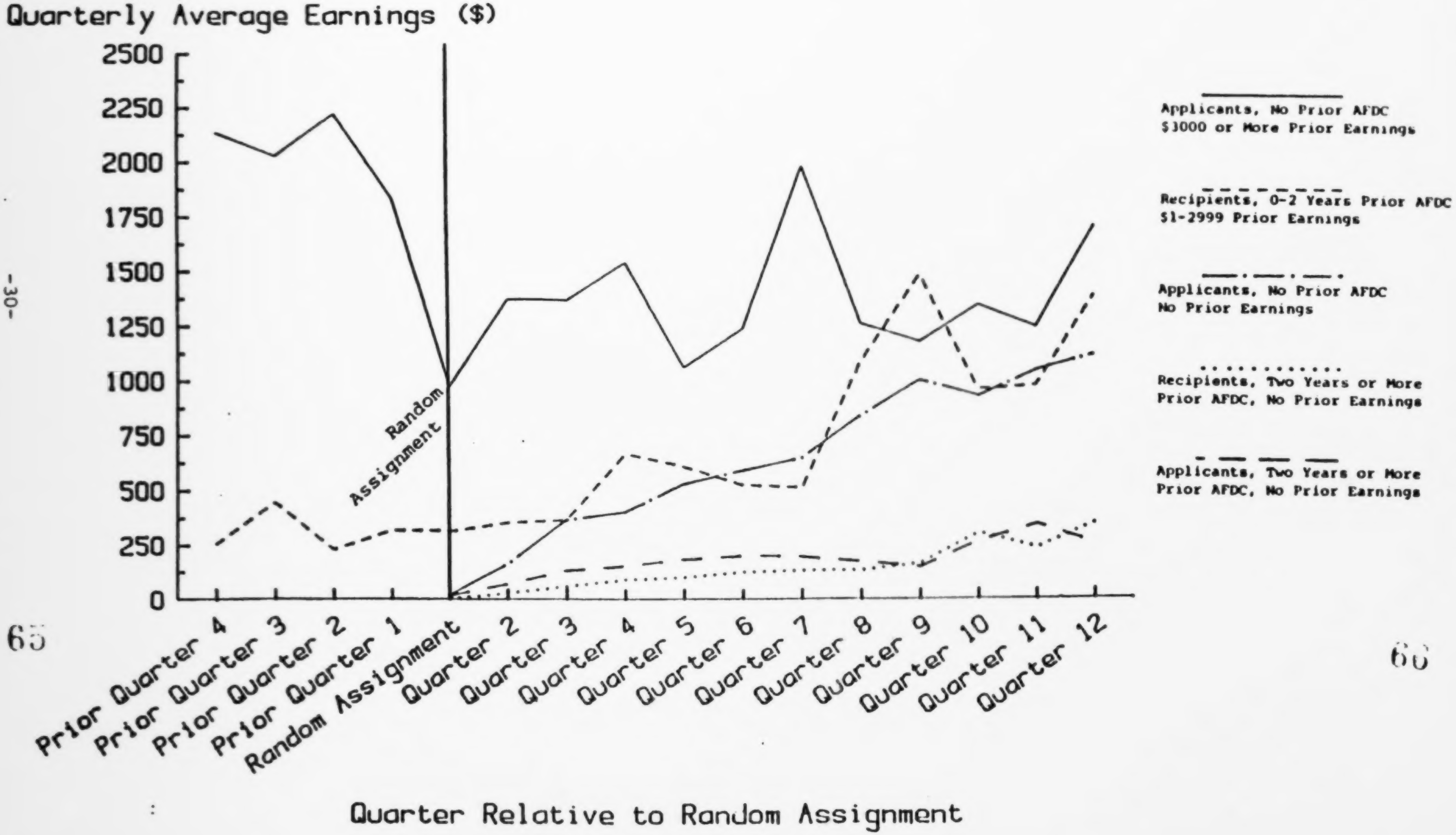
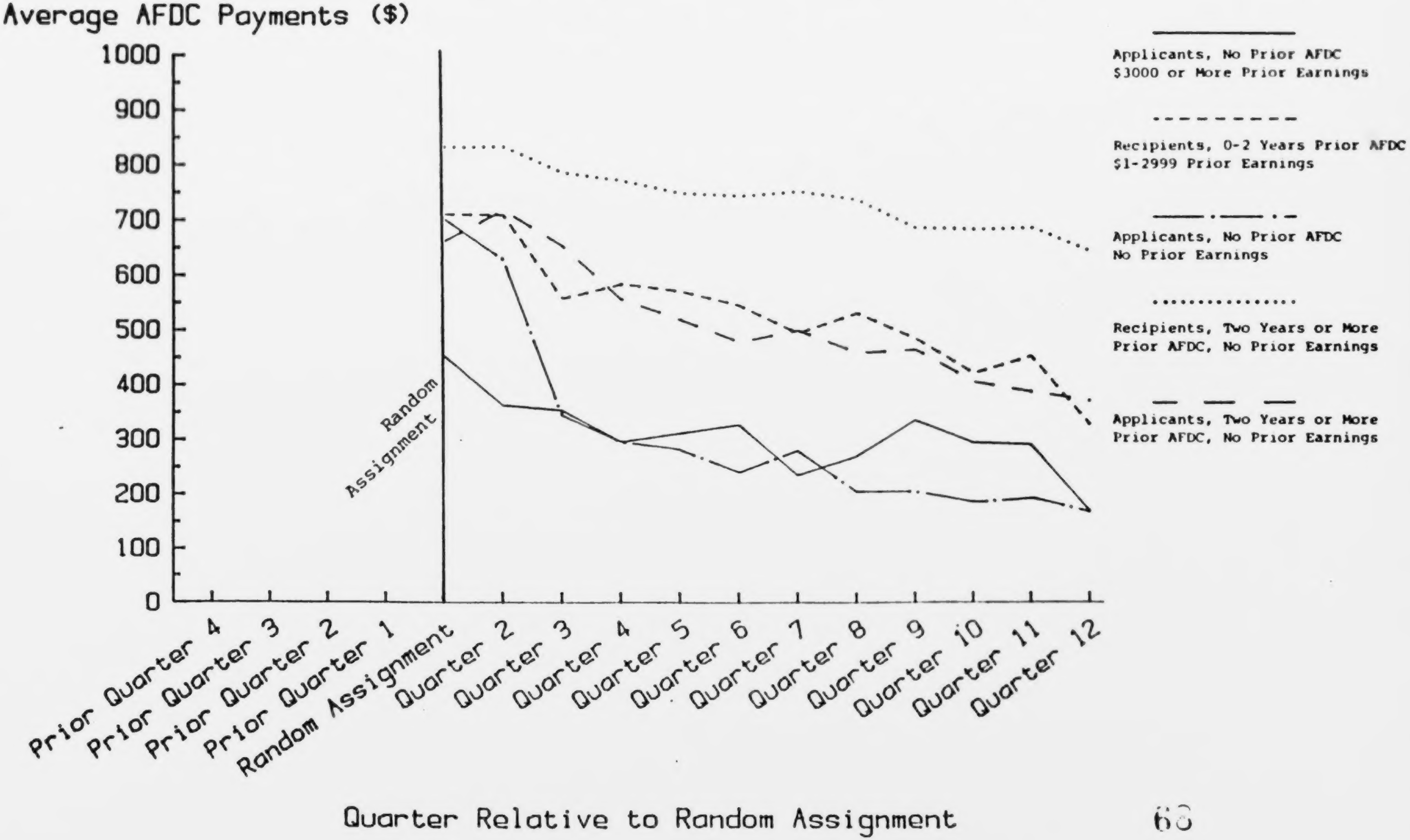


FIGURE 2.2
BALTIMORE CONTROLS, EARLY COHORT: QUARTERLY AVERAGE AFDC PAYMENTS, BY SUBGROUP



to $1,800 per quarter range over the follow-up period (these estimates count zero earnings for persons not employed). During the same period, subgroups with no recent employment history and a pattern of AFDC receipt for more than two years never reached average earnings over $400 per quarter.

The pattern of differences can change over time. In the case shown, subgroups at the top and the two subgroups at the bottom remained relatively stable, but the groups in between showed upward trends.

Differences in welfare payments across subgroups are also large. All subgroups show welfare receipt declining over time. But at the end of the 12 follow-up quarters, those recipients with two or more years of welfare receipt and no pre-program earnings were receiving three to four times the quarterly benefit payments of first-time applicants.

As can be seen, the majority of welfare receipt is accounted for by a minority of AFDC women. For example, in the Baltimore control sample, applicants constituted about one-half the sample but were consuming less than one-third of the AFDC expenditures on the control group at the three-year mark. At the other end of the spectrum, recipients of more than two years standing who had no earnings in the pre-program year were only one-third of the sample, but were receiving nearly half of the welfare payments to the control sample at the end of three years.[6] A further breakdown of recipients by whether or not they had high school diplomas revealed that dropouts who had not obtained even a GED comprised 18 percent of the sample but received 28 percent of the AFDC dollars.

IV. Subgroup Differences without Intervention

There are two complementary approaches to defining and studying subgroups. The first is to disaggregate the sample by specific individual characteristics, as is done in Table 2.2, and then define persons with certain combinations of characteristics as constituting a subgroup (such as recipients with no pre-program earnings and no high school diploma). Analysis of subgroups defined in this manner may provide considerable information about the potential results of highly specific targeting possibilities. These can be loosely linked to more general concepts of job readiness or welfare dependency. This study defines job ready as likely to become employed during the follow-up period. It defines dependent as likely to remain on AFDC during follow-up. The alternative approach uses all the individual characteristics in a multiple regression framework to predict job readiness and dependency. Subgroups are then categorized on the basis of similar predicted job readiness or dependency.

This report emphasizes the first approach. Table 2.4 shows differences in employment and welfare receipt rates during follow-up for several important subgroups of this study, pooling the data for controls in San Diego, Baltimore, Virginia, and Arkansas. These rates for controls again depict normal behavior (i.e., in the absence of the special program being evaluated). Quarter 6 was chosen as the end point because it is the last quarter for which all sample members had data. Data for Cook County controls were not included, because subgroup definitions for the impact sample for that program are not precisely comparable to the definitions for the other programs. The table shows the average levels of employment and welfare receipt of persons with a particular characteristic.

TABLE 2.4

AFDC CONTROLS: EMPLOYMENT AND WELFARE RECEIPT BY MAJOR SUBGROUP

Subgroup Characteristic	Average Quarterly Employment Rate, Quarters 4-6 (%)	Percent Receiving Any AFDC Payment, Quarter 6 (%)
Welfare Status		
Applicant	37.0	41.6
Recipient	22.7	72.4
Prior Year Earnings		
$3000 or More	60.0	38.5
$1-2999	43.8	49.3
None	17.2	60.4
Had Own AFDC Case		
Never	36.3	34.8
Two Years or Less	36.5	46.4
More Than Two Years	25.7	67.1
High School Diploma		
Yes	38.0	47.1
No	24.9	60.2
Child 12 or Under		
No	31.3	43.3
Yes	31.4	56.4
Number of Own Children		
One	32.7	51.7
More Than One	30.2	55.5
Currently Married		
Yes	32.3	47.4
No	30.7	58.0
Ever Married		
Yes	32.6	47.0
No	29.1	66.3
Age		
30 or Over	32.2	49.6
Under 30	30.2	59.4
Ethnicity[a]		
White	33.6	39.2
Black	29.4	62.4
Hispanic	43.9	42.1

SOURCE: MDRC calculations from MDRC Client Information Sheets; from the County of San Diego welfare records and Unemployment Insurance records from the EPP Information System; from the State of Maryland welfare and Unemployment Insurance records; from the Commonwealth of Virginia Unemployment Insurance earnings records, welfare records from the Virginia Automated Client Information System, and Fairfax County AFDC case files; and from State of Arkansas welfare and Unemployment Insurance records.

NOTES: Table entries were estimated from control samples pooled across San Diego, Baltimore, Virginia, and Arkansas. Estimates are not regression adjusted.

[a] For Baltimore, Virginia, and Arkansas the category "black" includes a small number of individuals in other non-white groups. In San Diego and Cook County, "white" includes a small number of non-black, non-hispanic, non-white persons.

It is important to note that these comparisons do not account for differences among individuals in any other characteristic. Regression analysis can take these differences into account. Table 2.5 shows the effects on employment and welfare receipt that are properly attributable to particular characteristics when differences attributable to the other characteristics have been accounted for.

It is useful to explain briefly how to interpret Table 2.5. For each characteristic or range of characteristics, the number shown in the table indicates the additional effect of the characteristic relative to some benchmark. Take the effect of being an AFDC recipient, for example. The fifth row down in the right-hand column indicates that being an AFDC applicant reduces by 16 percentage points the likelihood of receiving an AFDC payment in quarter 6, relative to the likelihood of a recipient receiving AFDC in quarter 6. The effect of prior earnings experience is interpretable as follows (see the seventh, eighth, and ninth entries in the left-hand column): Earning $3,000 or more in the year prior to random assignment adds 38.8 percentage points to the likelihood of being employed in quarter 6, relative to the likelihood of being employed in quarter 6 for someone with no earnings in the year prior to random assignment. Earning between $1 and $2,999 adds 23.8 percentage points to the likelihood of being employed in quarter 6, also relative to the likelihood for someone having no earnings in the year prior to random assignment. The increased likelihood of being employed in quarter 6 for those with higher earnings versus those with lower but still positive earnings in the prior year is 15 percentage points (38.8 minus 23.8).

As is clear from the tables, in these samples of adult enrollees,

TABLE 2.5

AFDC CONTROLS: EFFECTS OF SUBGROUP CHARACTERISTICS ON EMPLOYMENT AND WELFARE RECEIPT

Subgroup Characteristic	Average Quarterly Employment Rate, Quarters 4-6 (%)	Percent Receiving Any AFDC Payment, Quarter 6 (%)
Program Location		
San Diego	- 3.4*	- 8.0***
Baltimore	---	---
Virginia	+ 3.1**	-15.6***
Arkansas	-12.2***	-13.0***
AFDC Status		
Applicant	+ 2.0	-16.0***
Recipient	---	---
Prior Year Earnings		
$3000 or More	+38.8***	-11.1***
$1-2999	+23.8***	- 5.8***
None	---	---
Had Own AFDC Case		
Never	---	---
Two Years or Less	- 3.1*	+ 7.0***
More Than Two Years	- 6.4***	+16.2***
High School Diploma or Equivalent		
Yes	+ 8.1***	- 8.8***
No	---	---
Youngest Child Age		
Over 12 Years	---	---
6-12	+ 0.4	+ 9.6***
Less Than 6[a]	+ 3.9*	+ 5.3**
Number of Own Children		
One	---	---
More Than One	+ 0.5	+ 2.1
Age		
30 or Over	---	---
25-30	+ 0.1	+ 0.7
Less Than 25	- 1.8	+ 5.8**
Ever Married		
Yes	---	---
No	+ 0.5	+ 7.2***
Ethnicity[b]		
White	---	---
Black	- 0.5	+13.5***
Hispanic	+ 9.0***	+ 8.3**
Constant	+17.8***	+48.9***
Unadjusted R^2	.2060	.1774

SOURCE: See Table 2.4.

NOTES: Table entries are coefficients from a regression run on control samples pooled across San Diego, Baltimore, Virginia, and Arkansas. Sample size is 3869. Regression coefficients are always estimated as differences relative to a reference category. The reference categories are indicated with dashes. For example, being an AFDC applicant leads to a quarterly average employment rate that is 2 percentage points higher than the rate for AFDC recipients.

TABLE 2.5 (continued)

[a] In all programs except Arkansas about 10 percent of the samples had children less than 6 years old. In Arkansas 50 percent had children less than 6; only the Arkansas subgroup is broken out for separate analysis of impacts.

[b] For Baltimore, Virginia, and Arkansas the category "black" includes a small number of individuals in other non-white groups. In San Diego and Cook County, "white" includes a small number of non-black, non-hispanic, non-white persons.

prior earnings were the best single predictor of future employment and earnings.[7] Likewise, status as an AFDC applicant or recipient and being on welfare for more than two years were the best predictors of future welfare receipt.[8] It is important to note that applicant/recipient differences in AFDC receipt are only partly explained by differences between the two groups in prior employment and welfare history. A difference in welfare receipt of 16 percentage points between applicants and recipients remains even after controlling for all other differences in characteristics. This is in part because substantial proportions of applicants are never approved (and therefore never go on welfare) and partly because approved applicants tend to leave the system faster than longer-term recipients.

Job readiness and dependency are both related to demographic characteristics, as is to be expected, but not in the same way. This also can be seen in Table 2.5. Age, marital status and age of youngest child are important determinants of welfare but not employment status.[9] The effect of a high school diploma is similar for both. Prior earnings is a much better predictor of employment status than welfare status. Being a recipient rather than an applicant is not related to employment status, and prior welfare history is much less strongly related to employment status than it is to welfare status.

The estimates of Table 2.5 can be used to assign to each sample member a predicted future employment or welfare receipt rate. These can serve as scores on separate scales of job readiness and dependency.

Because combinations of low recent earnings and long welfare history may yield significantly more information than either alone, several single-trait and combination subgroups were also analyzed. Table 2.6 displays the

TABLE 2.6

AFDC CONTROLS: NORMAL EMPLOYMENT AND WELFARE BEHAVIOR FOR COMBINATIONS OF THE MAJOR SUBGROUPS

Subgroup	Percent of Applicants	Percent of Recipients	Average Quarterly Employment Rate, Quarters 4-6 (%)	Percent Receiving Any AFDC Payment, Quarter 6 (%)
First Tier				
Applicants With No Prior AFDC	31.2	---	37.0	31.2
Applicants With $3000 or More Prior Earnings	27.6	---	60.3	35.9
Second Tier				
Applicant Returnees	68.8	---	37.0	46.4
Applicants With Less Than $3000 Prior Earnings	72.4	---	28.1	43.8
Applicant Returnees With Less Than $3000 Prior Earnings	49.3	---	28.1	49.4
Third Tier				
All Recipients	---	100.0	22.7	72.4
Recipients With More Than Two Years on AFDC	---	72.9	19.6	76.4
Recipients With No Prior Earnings	---	75.8	14.6	75.7
Recipients With No Prior Earnings and More Than Two Years on AFDC	---	58.3	12.7	79.6

SOURCE AND NOTES: See Table 2.4.

subgroups that were found important in the subgroup analysis of impacts, organized into three categories, with some overlap among them:

First tier

Applicants with no prior AFDC

Applicants with $3,000 or more prior-year earnings

Second tier

Applicants who have been on AFDC before and have returned to the rolls (i.e., applicant returnees)

Applicants with less than $3,000 in prior-year earnings

Applicant returnees with less than $3,000 in prior-year earnings

Third tier

All recipients

Recipients with more than two years on AFDC

Recipients with no prior-year earnings

Recipients with no prior-year earnings and more than two years on AFDC

These three tiers correspond loosely to decreasing levels of job readiness and increasing levels of dependency, even though, as can be seen from the table, the two concepts do not yield exactly the same rankings. Subgroups within tiers overlap, and there is some overlap between the first and second tiers. The ranking is useful, nevertheless, as a way to summarize how the effectiveness of the services used in the five programs included in the study varies across levels of dependency (see Chapter 4, Section III).

V. Program Costs

In developing welfare employment policy, program impacts on employ-

77

ment, earnings, welfare receipt and other outcomes must be weighed against program costs. Cost differences by subgroup were calculated for the San Diego and Baltimore programs. This section briefly describes these cost differences and discusses implications for the overall subgroup analysis. A more detailed discussion of costs, together with an assessment of the benefit-cost implications of the subgroup impact and cost differences (i.e., the costs of the program over and above the costs of employment-related activities undertaken by controls) is available from MDRC.

Table 2.7 presents gross program costs, expressed on a per experimental basis, for the San Diego and Baltimore programs. The figures include the costs of serving nonparticipants as well as participants in the experimental groups, and are broken down by major program component. They are also disaggregated for the two major subgroups based on prior earnings and welfare experience.

Prior AFDC receipt was the most important characteristic associated with higher costs in both programs. The group with the longest welfare history, more than two years, had the highest costs. People in this subgroup stayed on welfare and in the programs longer and, in Baltimore, were assigned to the expensive services more often. Thus, for example, costs for recipients with more than a two-year welfare history were over 40 percent higher than for first-time applicants. In addition, in the Baltimore program, individuals without a high school diploma received more of the costly remedial education services than enrollees who already had a diploma.

Thus, subgroup costs did vary and the expenditures were higher for the more dependent. This pattern was probably present in all five programs,[10]

TABLE 2.7

SAN DIEGO AND BALTIMORE

PROGRAM COSTS PER EXPERIMENTAL, BY PROGRAM AND MAJOR SUBGROUP

Subgroup	Total Average Cost	Group Job Search	Work Experience	Other Program Activities	Support Services
San Diego Applicants	$786	$560	$91	$96	$38
Prior Year Earnings					
$3000 or More	843	627**	81	96	40
$1-2999	733	522	81	96	35
None	775	537	103	96	39
Had Own AFDC Case					
Never	729	534	66***	96	33
Two Years or Less	794	567	91	96	40
More Than Two Years	845	585	124	96	41
Baltimore Applicants[a]	843	173	51	329	195
Prior Year Earnings					
$3000 or More	702*	134	37*	294	150*
$1-2999	949	204	50	376	218
None	879	183	63	323	214
Had Own AFDC Case					
Never	804***	166	28***	342***	177***
Two Years or Less	694	146	48	251	159
More Than Two Years	1037	209	70	408	247
Baltimore Recipients[a]	1065	188	89	386	288
Prior Year Earnings					
$3000 or More	831	192	55	344	159
$1-2999	1041	215	82	363	267
None	1088	180	93	396	303
Had Own AFDC Case					
Never	635***	126	54*	213*	160**
Two Years or Less	862	163	70	315	214
More Than Two Years	1156	200	97	420	320

SOURCE: MDRC calculations from program cost and enrollment data (see Long and Caspar, 1987).

NOTES: Estimates are total costs incurred for experimentals and are averaged over participants and non-participants. F-tests were performed on variations in cost in each column for each subgroup dimension. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; *** = 1 percent.

[a] The cost components listed for Baltimore do not include the costs of sanctioning, and thus do not sum to total cost.

although it was most pronounced in Baltimore. Chapter 4 examines whether these greater expenditures produced greater impacts on the more dependent subgroups.

As that chapter makes clear, the overall subgroup variation in cost was small compared to the variation in impacts among subgroups. This was particularly true in San Diego, which had the same relatively short treatment sequence for all enrollees, and which did not include education and training. Decisions about alternative targeting strategies, therefore, hinge primarily on subgroup impact differences rather than cost differences.

Before the discussion of subgroup impacts, Chapter 3 presents a brief discussion of the methodology used for the analysis.

CHAPTER 3

MEASURING SUBGROUP IMPACTS AND ASSESSING PERFORMANCE INDICATORS

This chapter reviews the principal elements of the experimental research design and the methodology used in this study. The discussion is meant as a general guide, although parts of it are inevitably somewhat technical.

I. Experimental Design

Any analysis of program impacts is based on a comparison between the observed outcomes of a program and what would have occurred without it. As explained in Chapter 1, program outcomes are relatively easy to observe. But estimating the program impact requires calculation of the difference between observed outcomes and what outcomes would have been in the absence of the program.

A classical experimental design is often the preferred way of obtaining the standard for comparison. In such designs, clients are assigned on a random basis to either the experimental group, members of which are eligible for program services, or to the control group, members of which are only eligible for the services available without the program. The average outcomes of experimentals eligible for the program minus the average outcomes of controls are the program impact estimates. These measure the program achievements over and above the normal job-finding and welfare patterns of the eligible population. Random assignment ensures that the two groups are the same in terms of measured and, more important, unmeasured characteristics, permitting unbiased estimates of program impacts.

To maintain this lack of bias in the impact estimates, no changes can be made in the research group designations after random assignment. "Experimentals" remain experimentals and "controls" remain controls. Therefore, experimentals who did not, for some reason, participate in the programs are still counted as part of the experimental group in the calculation of impacts. Program impacts, therefore, are expressed on a per experimental rather than a per participant basis. The definition of subgroups adheres to the same principle. Subgroups are defined by pre-existing characteristics observed at enrollment, not by any subsequent behavior or activity.

II. Data Sources

Earnings and welfare data were assembled from administrative records. The use of such records offers several advantages. First, administrative records can be much less expensive than survey data, in part because registrants do not have to be recontacted during the follow-up. Records may also be more accurate than survey data because they do not depend on client recall of dollar amounts of earnings or welfare payments. Different rates of response by the experimental versus the control group -- often a source of bias in survey data -- are also not expected with records data.

Administrative records are, however, limited in their comprehensiveness and coverage. For example, quarterly earnings information can be obtained from the Unemployment Insurance (UI) system, but data on weeks worked, wages and hours worked are not available. Moreover, the information can only be obtained with a lag, and some delinquency in filing earnings reports on the part of employers is common in wage-reporting states. Another drawback is that state UI systems do not normally record the earnings of people who commute to work across state lines or uncovered employment. Given random assignment, however, none of these factors should affect experimental and control group outcomes differently.

In addition, administrative records in this study contain no information on people other than the research sample members. They do not, for example, provide the earnings of other family members, whose income (both earned and unearned) will affect a household's welfare dependency and general well-being.

The completeness and accuracy of the records data collected in this study were examined by comparing a small sample of data from the analysis tapes to the original paper or microfilm documents in state or county offices. Earnings and welfare payments were well matched. Further, a comparison of records and survey data from the Louisville WIN Laboratory and an earlier San Diego study suggests that the two sources yielded relatively similar information although the survey self-reports somewhat understated welfare receipt.[1]

Records data were merged with demographic and program activity information to form a single program data base, with a new record compiled for each sample member. Each record contains the client's employment background and welfare history in addition to aK series of outcome measures (quarterly UI earnings, monthly AFDC payments) running from the point of entry into the sample (i.e., the date of random assignment) through to the end of the followup. Program activities and dates are also included. The earlier a person entered the sample, the more follow-up data are available.

No sample member has less than six quarters of earnings data and 18 months of welfare data. This is the minimum available in San Diego and Cook County. At least ten quarters are available in Baltimore and Virginia, twelve in Arkansas.

The major data sources for all the programs analysed are summarized below:[2]

> Client Information Sheets, one-page questionnaires filled out by client and staff as part of the random assignment process, provide information on the demographic characteristics of sample members. All principal subgroups, with the exception of the subgroups identified by prior earnings, were defined using this information.
>
> State Unemployment Insurance (UI) Earnings Records provide quarterly employment and earnings data reported by employers for each calendar quarter: e.g., January, February and March; April, May and June.
>
> AFDC records supply information on monthly AFDC (i.e., welfare) grants. Monthly AFDC data are grouped by three-month periods, where the first month of the first quarter of follow-up is the month of enrollment.
>
> Unemployment Insurance Benefit Records supply information on monthly UI benefit payments.
>
> Program Activity records provide information on program services, participation and deregistration.

It is important to note that Client Information Sheets were not available for the full impact sample in Cook County. For this reason, fewer subgroups could be defined there, and those are typically not directly comparable to the subgroups for the other four programs.

Since random assignment can occur in the first, second, or third month of a calendar quarter, the first quarter of UI earnings can contain pre-program earnings for some sample members. The first quarter of earnings is therefore not considered a follow-up quarter in the impact analysis and is omitted from cumulative estimates of program impacts.

III. Choice of Follow-up Period

MDRC's research to date has shown certain patterns of outcomes for experimentals and controls over time. Typically, the outcomes for experimentals and controls were similar in the quarter of random assignment but began to differ in quarter 2, even though many experimentals did not join activities for as long as six months after enrollment. The experimental-control differences grew slowly, with the difference often peaking at the one-year point or beyond.

This report divides follow-up into an immediate post-random assignment period (quarters 1 through 3) and a longer-term follow-up period (quarters 4 and following). Quarters were averaged -- which helps to eliminate some of the transitory quarter-to-quarter variation in earnings. Earnings, as well as employment, AFDC receipt and payment amounts, are expressed as quarterly averages per person. Averages for the immediate and longer-term outcomes were calculated separately. It should be emphasized that the longer-term averages contain more quarters of data for persons who entered the samples early. This averaging procedure has the disadvantage that it does not explicitly estimate quarter-by-quarter time trends in impacts.

The longer-term follow-up period was selected as the focus of this subgroup analysis because subgroup differences appearing in the later quarters are the best indicators of long-run effects and are therefore likely to be more indicative of the total impact differences among subgroups. The training activities and education programs in Baltimore, which take as long as one year to complete, require a long follow-up period, making it even more important to focus on the later periods.

Statistical tests of significance are reported for differences between experimentals and controls within subgroups. The differences between impacts for pairs of subgroups were also tested; the results of these tests, which were not often statistically significant, are omitted from the tables but are occasionally mentioned where appropriate. Some of these tests, along with other special statistical considerations relevant to the empirical comparisons among subgroups, are discussed in Appendix A.

IV. The Subgroup Impact Regression Model

A simple difference between average outcomes for experimental and control groups is sufficient to estimate impacts reliably in a carefully implemented experimental design. Use of linear regression lends extra precision to the estimates and corrects for minor differences in pre-program characteristics between experimentals and controls. For this reason, the estimates reported in this paper are regression-adjusted.

In addition, regression techniques have been used to produce two sets of subgroup impacts. The first set takes the point of view of the program administrator who asks: "Can I improve efficiency by targeting services to registrants with a single subgroup characteristic?" For example, it may be useful to find out if sample members with a high school diploma have different impacts than those without diplomas, ignoring differences in any other demographic characteristics (the kind of estimates shown in Chapter 2, Table 2.4). These impact estimates are unconditional estimates, and are the focus of Chapter 4. Such subgroup estimates do not take into account impact differences associated with other demographic and background characteristics. For example, women without a high school diploma generally have a weaker work record, but unconditional estimates do not explain what part of the diploma effect is due to the work history characteristic itself rather than other characteristics of individuals with weak work records. Regression, in this case, only serves the purpose of increasing precision and adjusting for minor pre-existing experimental-control differences.

Two or more characteristics can be included in unconditional estimation as *interactions*, and these are often useful to program operators. To continue the example above, the sample may be split four ways: persons with and without diploma, further divided by employed/not employed in the recent pre-program period. Impacts calculated for each of these four subgroups may help to establish whether it is worthwhile to target services to a narrow subgroup defined by diploma and prior employment status. This approach provides information about targeting on the basis of two subgroup characteristics, without controlling for other factors.

Regression analysis can be used to generate *conditional* estimates. These estimates hold all subgroup characteristics constant except the one in question. That is, any conditional impact difference associated with a high school diploma would indicate the importance of the schooling credential itself, eliminating effects due to prior employment record and other characteristics. If conditioning on prior employment status nullified the diploma effect, then the prior-employment difference across diploma subgroups would be the "real" reason for the diploma impact.[3]

Both unconditional and conditional estimates are important, depending on the questions asked. Unconditional estimates are presented and discussed in the next chapter because they address questions of targeting with limited information. Conditional estimates, however, are required for the testing of performance measures in Chapter 5. Conditional estimates will be discussed in Chapter 4 only insofar as they raise issues regarding the conclusions drawn from the unconditional estimates.

V. Testing Performance Indicators

A handful of prior studies have attempted to test the correlation between various measures of performance and net program impact. These studies generally did not have experimental comparison data, but their techniques are similar to the ones used in this study of performance measures.

The basic approach is as follows:

1. Obtain an estimate of net program impact for each individual in the treatment group;
2. Create a measure of program performance -- e.g., did the sample member enter employment, what were his/her wages?
3. Compute correlation coefficients between the net impact and the performance measures, with measures with the greatest correlation being identified as the "best" performance indicators;
4. As a supplemental analysis, determine whether two indicators work better than one. Compute a regression of net impact on two performance indicators and report the coefficients and their statistical significance. In this way, it may be possible to determine that one indicator has more power than another or is a useful supplement.

This procedure has remained approximately the same since studies in the mid-1970s correlated performance measures with the impacts of certain pre-CETA employment programs.[4]

The difficult part of this process is the first step: the estimation of a net impact for each individual.[5]

Early studies of performance indicators estimated individual-level impacts without experimental data, and thus had to depend on impact estimates from participant/nonparticipant comparisons adjusted by regression for various demographic and participation variables, such as type of treatment and length of stay. Thus, while these studies have used essentially the same procedure to estimate individual impacts as used in the random assignment evaluations, the estimates they generated will be biased to the extent that the regression models used were not able to control adequately for differences between the participant and nonparticipant groups.

Interpretation of the correlations can be problematic, because they apply to the programs as implemented when the reason for testing performance indicators may be to establish standards that will change the way that programs are implemented. For this reason it is important, in assessing the validity of any particular class of indicators, to consider other evidence as well.

CHAPTER 4

SUBGROUP DIFFERENCES IN IMPACTS

This chapter summarizes program impact differences for subgroups of the WIN-mandatory AFDC caseload in the five work/welfare areas included in this report -- San Diego, Baltimore, several counties in Virginia, part of Little Rock and one less urbanized county in Arkansas, and Cook County in Illinois. The main focus is on the subgroups of WIN-mandatory registrants listed in Chapter 2, defined according to prior welfare and employment experience. Defining the subgroups along the straightforward dimensions of prior welfare and earnings history makes the conclusions on targeting of direct use to program operators, since the subgroups can be readily identified for a variety of eligible populations.

One thrust of the findings is that the least dependent subgroups (e.g., those with prior-year earnings above $3,000 or no previous welfare receipt) generally experienced below-average program impacts, that were rarely statistically significant and often the smallest estimates of any major subgroup. These findings suggest that a policy of targeting low-cost program components only to those in the WIN-mandatory caseload who are most "job ready" is not efficient. Although such a policy may result in a high "entered employment" rate among participants or a high rate of AFDC case closures, administrators cannot be confident that such an approach will lead to real impacts on earnings or income.

Above average and statistically significant earnings impacts occurred most consistently for a subgroup in the middle range of dependency (applicants who had already had at least one spell on AFDC). The earnings impacts on the most dependent subgroups (e.g., those with no prior-year earnings who are welfare recipients and have prior welfare histories) were typically below average, and did not occur consistently across programs. This suggests that focusing solely on these groups may also not be the most effective strategy, at least in the context of the mandatory, mass participation, relatively low-cost programs that are the focus of this study.

There were no impacts on welfare payments for the least dependent groups. There were modest impacts on welfare receipt for groups in the mid- and high-dependency range, often not statistically significant, however.

Impact differences among subgroups defined according to characteristics other than prior work and welfare generally were not consistent across programs. The best predictor among these other characteristics was marital status. The bulk of welfare savings in the four programs for which there are data on marital status came from the not-married and never-married subgroups.

The tables report tests for the statistical significance of impacts for the individual subgroups. Significance tests across subgroups were also performed. Estimates of impacts for one subgroup may be described as "larger" or "smaller" than those of another, but without passing the statistical criteria these relationships lack the high level of certainty generally applied in social science research. Subgroups which appear to have larger impact estimates may be subjected to cross-subgroup statistical tests of confidence. Such tests were not usually statistically significant, but where they contributed to confidence in the estimated differences in impacts across subgroups they are noted. For example, earnings gains for applicant returnees were statistically significantly greater than those for new applicants and all recipients combined (i.e., the balance of the sample) for two of the three programs where such comparisons were possible. Selected tests of this sort, together with a discussion of other statistical issues involved in subgroup comparisons, appear in Appendix A.

In each table in this chapter the subgroups are shown with the most employable and/or least dependent subgroup at the top and the least employable and/or most dependent at the bottom. It should be noted that not all subgroups could be defined for all programs.

This chapter is organized in three sections. The first section analyzes impacts for the major subgroups and focuses on cross-program comparisons. The second examines other results of interest within each of the programs studied. Since the patterns that emerge are not completely uniform across programs, across subgroups or across dependency measures, the discussion is inevitably somewhat complex. The final section of the chapter provides a summary of the major findings.

I. The Major Subgroups

One way to judge the magnitude of impacts for a subgroup is to compare the estimates with those for another subgroup. Are they higher or lower? Is the subgroup estimate larger or smaller than the full sample estimate? Another straightforward way is to classify impacts as falling above or below some cutoff value. In this chapter, employment and welfare cutoffs were selected so that impact estimates for about half the subgroups would fall above and half below the cutoff values. Impacts estimated to be above the cutoff values are called above average, and those below are labelled below average. The cutoffs chosen are, however, technically more akin to the concept of a median than a mean. The choice of cutoffs is to some degree arbitrary; they could be lower or higher depending on the stringency needed.

The cutoff criteria define above average impacts for each outcome measure, respectively, as:

- employment rate increases of more than 4 percentage points a quarter
- earnings increases of more than $100 a quarter
- welfare receipt reductions of more than 2 percentage points a quarter
- welfare savings of more than $20 a quarter.

Estimates that fell below the cutoff on any of these definitions were considered below average even if statistically significant. Estimates that were above the cutoff but not statistically significant were considered uncertain. Some degree of arbitrariness is inherent in these cutoffs, since typical earnings rates and AFDC payment schedules differ across states and localities. It should also be emphasized again that program impacts are only one factor in targeting; other factors are also important, depending on program objectives.

The major subgroups used in the analysis are defined along three dimensions: whether they were applicants or recipients when they entered the study sample, whether they worked and how much they earned in the year before sample entry, and whether and how long they had been on welfare at sample entry.

The minimum length of follow-up for earnings and AFDC payment data is six quarters in San Diego and Cook County, ten quarters in Baltimore and Virginia, and 12 quarters in Arkansas.

The most important definitional problem for subgroups concerns the distinction between applicants and recipients in Cook County. Since this information was not recorded explicitly for the research sample the distinction had to be approximated. Sample members who received no AFDC payment in the three months prior to random assignment were classified as applicants; all others were classified as recipients. This procedure would have yielded groups that were relatively comparable to the groups in other states except that, in Cook County, only applicants whose grants had already been approved were enrolled in the program. The applicant samples in the other states included all applicants, whether their grants were subsequently approved or not. The significance of this for the research is that the Cook County applicant subgroups had higher welfare receipt during the follow-up period than did the other applicant subgroups.

A. Welfare Status

The evidence suggests that employment and earnings effects are larger for applicants than for recipients, although the evidence is not strong and is not completely consistent across programs (Table 4.1). Employment and earnings impacts for applicants were larger than those for recipients in two of the four programs that served both categories, although the increment was statistically significant in only one sample. Estimates for the two groups were similar in a third program (Virginia). Employment increases were over 4 percentage points for four of the applicant samples but only for one recipient sample. At the same time, earnings gains topped

TABLE 4.1

AFDC APPLICANTS AND RECIPIENTS: UNCONDITIONAL IMPACTS ON
EMPLOYMENT, EARNINGS, AFDC INCIDENCE AND PAYMENTS,
BY PROGRAM AND WELFARE STATUS

Subgroup, Welfare Status, and Program	Percent of Sample[a]	Percent Employed Quarterly Quarters 4 - Last (%)			Average Earnings Per Quarter Quarters 4 - Last ($)		
		Experimental	Control	Difference	Experimental	Control	Difference
Full Sample							
San Diego[b]	100.0	41.9	37.4	+4.5***	891	773	+118**
Baltimore	100.0	38.8	35.3	+3.5***	730	634	+ 96***
Virginia	100.0	37.6	33.5	+4.1***	613	541	+ 72**
Arkansas	100.0	23.5	18.1	+5.4***	327	257	+ 70**
Cook County	100.0	23.9	23.0	+0.9	470	451	+ 19
Applicants							
San Diego	100.0	41.9	37.4	+4.5***	891	773	+118**
Baltimore	50.1	46.5	42.2	+4.3**	997	825	+172***
Virginia	40.3	47.5	43.2	+4.3**	819	738	+ 80
Arkansas	59.4	30.7	23.3	+7.4***	449	341	+107**
Cook County[c]	33.7	31.7	32.8	-1.1	693	731	- 38
Recipients							
Baltimore	49.9	31.1	28.3	+2.8	472	436	+37
Virginia	59.7	30.9	26.8	+4.1**	474	406	+69*
Arkansas	40.6	13.1	10.5	+2.5	150	131	+19
Cook County[c]	66.3	19.9	17.9	+1.9**	350	304	+46**

Subgroup, Welfare Status, and Program	Percent of Sample[a]	Percent Receiving AFDC Monthly Quarters 4 - Last (%)			Average AFDC Payments Per Quarter Quarters 4 - Last ($)		
		Experimental	Control	Difference	Experimental	Control	Difference
Full Sample							
San Diego[b]	100.0	32.3	34.0	-1.7	436	469	-33
Baltimore	100.0	56.5	57.7	-1.3	496	501	- 5
Virginia	100.0	40.0	41.8	-1.8	322	345	-23*
Arkansas	100.0	36.4	43.4	-7.0***	192	232	-40***
Cook County	100.0	68.6	70.5	-1.9**	633	646	-13
Applicants							
San Diego	100.0	32.3	34.0	-1.7	436	469	-33
Baltimore	50.1	43.0	45.4	-2.4	366	380	-14
Virginia	40.3	23.5	26.2	-2.6	190	210	-19
Arkansas	59.4	24.9	29.7	-4.7*	130	156	-26*
Cook County[c]	33.7	54.5	56.9	-2.3	470	485	-15
Recipients							
Baltimore	49.9	70.2	70.2	+ 0.0	627	622	+ 5
Virginia	59.7	51.2	52.1	- 0.9	412	436	-24
Arkansas	40.6	53.0	63.6	-10.6***	283	344	-60***
Cook County[c]	66.3	75.9	77.4	-1.5	731	744	-13

TABLE 4.1 (continued)

SOURCE: MDRC calculations from the County of San Diego welfare records and Unemployment Insurance records from the EPF Information System; from the State of Maryland welfare and Unemployment Insurance records; from the Commonwealth of Virginia Unemployment Insurance earnings records, welfare records from the Virginia Automated Client Information System, and Fairfax County AFDC case files; from the State of Arkansas welfare and Unemployment Insurance records; and from the Illinois Department of Public Aid AAID system records and the Illinois Unemployment Insurance system earnings records.

NOTES: These data are regression-adjusted using ordinary least squares, controlling for pre-random assignment characteristics of sample members. Dollar-denominated estimates include zero values for sample members not employed or for sample members not receiving welfare. Estimates for applicants and recipients were obtained from separate regressions for each program. There may be some discrepancies in calculating sums and differences due to rounding.

Sample Sizes are as follows:

	Applicants	Recipients
San Diego	2381	---
Baltimore	1380	1377
Virginia	1269	1881
Arkansas	670	457
Cook County	2668	5187

Samples for San Diego and Illinois exclude a second experimental group, not analyzed in this report.

A two-tailed t-test was applied to differences between experimental and control groups. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent, *** = 1 percent.

[a] Percent of full sample.

[b] San Diego served only applicants.

[c] The definitions of "applicant" and "recipient" for Cook County differ from the other programs. See text for discussion.

$100 per quarter for three of the five applicant samples but none of the recipient samples. Averaging across programs (giving each program equal weight) yields earnings gains that were about twice as large for applicants as for recipients.

A supplemental analysis using regression analysis to control for demographic differences across sample members indicated that the impact differences did indeed stem from the applicant/recipient distinction, and not from other factors.[1]

Impacts on welfare do not follow a clear subgroup pattern. Reductions in the percent receiving AFDC were over 2 percentage points for four of the five applicant samples versus only one of the four recipient samples, but typically were not statistically significant. The only statistically significant impacts were in Arkansas, where the reduction was larger for recipients than for applicants. Impacts on AFDC payments for applicants versus recipients show no clear pattern.

B. Prior Earnings

Among applicants, the prior earnings results indicate that the most employable groups did not have the largest impacts (see Table 4.2). Only one program (Virginia) achieved its maximum employment and earnings increases for applicants with $3,000 or more in prior-year earnings. Two programs (Baltimore and Arkansas) showed maximum impact for the subgroup with prior-year earnings in the $1 - $2,999 range. San Diego showed the largest gains among applicants with zero earnings in the prior year.

Among recipients, since very few people had prior earnings in the top earnings category, all individuals with prior-year earnings were grouped together. For recipients with some prior-year earnings only one of the

TABLE 4.2

AFDC APPLICANTS AND RECIPIENTS: UNCONDITIONAL IMPACTS ON EMPLOYMENT, EARNINGS, AFDC INCIDENCE AND PAYMENTS, BY PROGRAM, WELFARE STATUS, AND PRIOR YEAR EARNINGS SUBGROUP

Subgroup, Welfare Status, and Program	Percent of Sample[a]	Percent Employed Quarterly Quarters 4 - Last (%)			Average Earnings Per Quarter Quarters 4 - Last ($)		
		Experimental	Control	Difference	Experimental	Control	Difference
Prior Year Earnings							
Applicant							
$3000 or More							
San Diego	28.8	60.7	58.7	+1.9	1444	1482	- 39
Baltimore	31.9	65.0	62.5	+2.5	1453	1435	+ 18
Virginia	29.2	65.6	56.1	+9.5**	1348	1041	+307***
Arkansas	12.5	60.8	53.4	+7.3	896	884	+ 12
Cook County[b]	33.0	53.9	52.2	+1.7	1373	1336	+ 37
$1-2999							
San Diego	22.9	42.9	41.7	+1.2	813	729	+ 84
Baltimore	29.3	51.5	45.0	+6.5*	1068	729	+339***
Virginia	28.4	52.4	48.2	+4.2	801	718	+ 83
Arkansas	22.5	56.9	43.3	+13.6***	795	598	+197*
Cook County[b]	20.4	33.5	38.4	-4.8	555	722	-167
None							
San Diego	48.4	30.3	22.8	+7.5***	601	375	+225***
Baltimore	38.8	27.7	23.6	+4.1	569	398	+171*
Virginia	42.3	31.7	30.9	+0.8	465	544	- 79
Arkansas	64.9	15.8	10.6	+5.3*	240	146	+ 95
Cook County[b]	46.5	15.0	16.5	-1.5	268	301	- 33
Recipients							
Some							
Baltimore	27.2	50.3	52.7	-2.4	799	970	-171**
Virginia	23.0	51.8	47.2	+4.5	848	783	+ 65
Arkansas	8.5	39.4	46.3	-6.9	486	578	- 92
Cook County[b]	19.8	42.2	38.0	+4.2**	865	685	+179***
None							
Baltimore	72.8	23.9	19.1	+4.7**	343	240	+104**
Virginia	77.0	24.7	20.8	+3.9**	363	293	+ 70
Arkansas	91.5	10.6	7.2	+3.4	118	89	+ 29
Cook County[b]	80.2	14.4	13.0	+1.3	223	211	+ 12

(continued)

TABLE 4.2 (continued)

Subgroup, Welfare Status, and Program	Percent of Sample[a]	Percent Receiving AFDC Monthly Quarters 4 - Last (%)			Average AFDC Payments Per Quarter Quarters 4 - Last ($)		
		Experimental	Control	Difference	Experimental	Control	Difference
Prior Year Earnings							
Applicants							
$3000 or More							
San Diego	28.8	26.5	25.9	+0.6	326	323	+ 3
Baltimore	31.9	36.5	34.8	+1.7	295	288	+ 7
Virginia	29.2	22.7	21.9	+0.8	182	174	+ 8
Arkansas	12.5	27.2	26.7	+0.5	138	137	+ 1
Cook County[b]	33.0	46.4	47.0	-0.6	406	403	+ 3
$1-2999							
San Diego	22.9	30.7	33.7	-2.9	409	471	-63
Baltimore	29.3	42.2	48.0	-5.8	367	405	-38
Virginia	28.4	24.8	28.2	-3.4	201	240	-39
Arkansas	22.5	25.2	33.6	-8.4	126	177	-51*
Cook County[b]	20.4	56.7	59.3	-2.6	460	493	-33
None							
San Diego	48.4	36.5	38.9	-2.4	514	554	-40
Baltimore	38.8	48.8	51.9	-3.2	424	437	-13
Virginia	42.3	23.3	27.8	-4.5	189	215	-26
Arkansas	64.9	24.4	28.9	-4.4	130	152	-22
Cook County[b]	46.5	59.4	62.8	-3.4	519	540	-21
Recipients							
Some							
Baltimore	27.2	62.5	58.8	+3.7	528	502	+26
Virginia	23.0	41.1	41.2	-0.1	326	342	-16
Arkansas	8.5	43.0	46.1	-3.1	206	241	-35
Cook County[b]	19.8	62.1	67.0	-4.8**	567	607	-40*
None							
Baltimore	72.8	73.1	74.4	-1.2	666	666	+ 1
Virginia	77.0	54.2	55.3	-1.1	437	464	-26
Arkansas	91.5	54.0	65.2	-11.3***	291	353	-63***
Cook County[b]	80.2	79.2	79.9	-0.6	772	778	- 6

SOURCE AND NOTES: See Table 4.1.

[a] Percent of applicants and percent of recipients.

[b] The definitions of "applicant" and "recipient" for Cook County differ from the other programs. See text for discusion.

four programs had above average effects on earnings. The great majority of recipients had no prior-year earnings. Only one program had an above average effect on this group although all the impacts were positive.

Welfare effects were again smaller than employment and earnings effects. It is of considerable interest to note that applicants with the best prior earnings records had no dollar welfare savings in any of the five programs. This was true even in Virginia, where this subgroup obtained above-average earnings increases. In contrast, all ten of the welfare receipt estimates for applicants with $1-2,999 or zero prior earnings (in the middle of the employability range) showed reductions greater than 2 percentage points, although none was statistically significant. All except one of these subgroups also showed above average dollar savings, though again not generally statistically significant. Only three of the reductions among recipients exceeded 2 percentage points, although some recipient groups did have dollar savings.

C. Prior Welfare

Impacts by length of prior welfare receipt are shown in Table 4.3. The Cook County sample is excluded, since information about length of prior welfare was not available.

With respect to employment and earnings, the subgroups at the top of the breakdown show no consistent increases. These first-time applicants are clearly the least dependent as a group, although some of them may have received welfare on their mother's grant as a child. Nor are any employment or earnings effects statistically significant for this subgroup. Only one employment estimate (Arkansas) is greater than a 4 percentage point gain; only one earnings impact (Baltimore) exceeds $100.

TABLE 4.3

AFDC APPLICANTS AND RECIPIENTS: UNCONDITIONAL IMPACTS ON EMPLOYMENT, EARNINGS, AFDC INCIDENCE AND PAYMENTS, BY PROGRAM, WELFARE STATUS, AND AFDC HISTORY SUBGROUPS

Subgroup, Welfare Status, and Program	Percent of Sample[a]	Percent Employed Quarterly Quarters 4 - Last (%)			Average Earnings Per Quarter Quarters 4 - Last ($)		
		Experimental	Control	Difference	Experimental	Control	Difference
Had Own AFDC Case							
Applicants							
Never							
San Diego	33.4	44.1	41.8	+2.3	1018	981	+ 37
Baltimore	22.7	46.6	47.1	-0.4	1136	1015	+121
Virginia	26.2	44.9	43.0	+1.9	868	881	- 13
Arkansas	56.3	27.3	22.4	+4.9	394	368	+ 26
Two Years or Less							
San Diego	38.7	41.4	35.3	+6.1**	898	732	+165**
Baltimore	41.8	51.7	45.9	+5.8*	1109	940	+169*
Virginia	31.7	50.1	43.2	+7.0*	888	719	+169*
Arkansas	36.7	35.8	25.6	+10.2**	541	329	+212**
More Than Two Years							
San Diego	27.9	40.0	35.2	+4.8	731	585	+146
Baltimore	35.5	40.6	35.0	+5.6*	776	568	+208**
Virginia	42.2	47.2	43.2	+3.9	736	663	+ 73
Arkansas	7.0	31.9	18.6	+13.3	402	194	+208
Recipients							
Two Years or Less[b]							
Baltimore	26.2	43.0	39.0	+4.0	753	638	+115
Virginia	28.2	34.8	35.0	-0.2	577	595	- 18
Arkansas	35.0	19.6	16.9	+2.7	241	214	+ 27
More Than Two Years							
Baltimore	73.8	26.9	24.5	+2.4	367	368	- 0
Virginia	71.8	29.5	23.6	+5.8***	436	327	+110**
Arkansas	65.0	9.6	7.1	+2.4	100	85	+ 14

(continued)

TABLE 4.3 (continued)

Subgroup, Welfare Status, and Program	Percent of Sample[a]	Percent Receiving AFDC Monthly Quarters 4 - Last (%)			Average AFDC Payments Per Quarter Quarters 4 - Last ($)		
		Experimental	Control	Difference	Experimental	Control	Difference
Had Own AFDC Case							
Applicants							
Never							
San Diego	33.4	22.7	23.0	-0.2	314	320	- 5
Baltimore	22.7	33.3	35.8	-2.5	287	295	- 9
Virginia	26.2	16.1	20.0	-3.8	132	160	-28
Arkansas	56.3	19.1	24.2	-5.1	96	127	-31
Two Years or Less							
San Diego	38.7	33.1	36.9	-3.7	436	510	-74*
Baltimore	41.8	40.2	40.9	-0.7	344	352	- 8
Virginia	31.7	20.3	23.7	-3.4	169	188	-19
Arkansas	36.7	33.7	35.7	-2.0	181	189	- 8
More Than Two Years							
San Diego	27.9	42.6	43.0	-0.4	580	589	- 8
Baltimore	35.5	52.3	56.7	-4.4	443	468	-25
Virginia	42.2	30.5	31.8	-1.3	243	257	-14
Arkansas	7.0	23.9	39.5	-15.6	129	204	-75
Recipients							
Two Years or Less[b]							
Baltimore	26.2	54.8	58.7	-3.9	483	508	-25
Virginia	28.2	41.0	37.7	+3.3*	324	305	+19
Arkansas	35.0	37.5	54.4	-16.9***	201	293	-92**
More Than Two Years							
Baltimore	73.8	75.7	74.2	+1.5	680	661	+19
Virginia	71.8	54.8	58.6	-3.8*	444	492	-48**
Arkansas	65.0	61.7	69.0	-7.3	329	373	-44*

SOURCE AND NOTES: See Table 4.1.

[a] Percent of applicants and percent of recipients.

[b] Includes a small number of recipients who reported never having had their own AFDC case.

Further down the table, applicants with a welfare history had employment and earnings impacts that were almost always above average and were more often than not statistically significant. Among these applicant returnees, the length of prior welfare history did not make much difference, although the impacts for those with two years or less were more likely to be statistically significant. Seven of the eight employment impacts for these two subgroups were over 4 percentage points; seven of the earnings impacts were over $100. If estimates were available for Cook County, they would probably not show impacts for these subgroups, since neither applicants as a whole nor any of the applicant subgroups that can be broken out showed much impact there. With respect to welfare savings, the only clear pattern was that in no program did first-time applicants -- the least dependent subgroups -- have the largest impacts.

D. Combination Subgroups

As shown in Chapter 2, the combination of low prior earnings and a long welfare history leads to longer periods of time on welfare than does either of those characteristics alone. It is therefore of interest to examine program impacts for different earnings/welfare history combinations.

Table 4.4 presents impact results for several pairs of such subgroups. Each pair consists of a less dependent and a more dependent subgroup. The combinations were chosen to split each of the applicant and recipient samples into two parts as equal in size as possible.[2] The more dependent of the two applicant subgroups was defined as follows:

TABLE 4.4

AFDC APPLICANTS AND RECIPIENTS: IMPACTS FOR SUBGROUPS COMBINING PRIOR EARNINGS, PRIOR AFDC RECEIPT, AND HIGH SCHOOL DIPLOMA STATUS

	Earnings Impact, Quarters 4 - Last ($)						
	Applicants				Recipients		
Subgroup	San Diego	Baltimore	Virginia	Arkansas	Baltimore	Virginia	Arkansas
Lower Prior Earnings Plus Higher Prior AFDC[a]							
No	+ 87	+ 86	+153*	+ 36	-48	+35	- 0
Yes	+151**	+253***	+ 20	+202**	+88	+94*	+28
Lower Prior Earnings Plus Higher Prior AFDC Plus No High School Diploma							
No	+109*	+174**	+ 75	+ 95*	+14	+63	+ 0
Yes	+158	+165	+ 95	+150	+57	+78	+43

	AFDC Payment Impact, Quarters 4 - Last ($)						
	Applicants				Recipients		
Subgroup	San Diego	Baltimore	Virginia	Arkansas	Baltimore	Virginia	Arkansas
Lower Prior Earnings Plus Higher Prior AFDC[a]							
No	- 3	- 8	- 9	-29	+18	+ 8	-83**
Yes	-63*	-19	-29	-22	- 1	-48**	-48*
Lower Prior Earnings Plus Higher Prior AFDC Plus No High School Diploma							
No	-26	-11	- 2	-18	+20	+ 4	-69**
Yes	-69	-19	-63*	-54*	-14	-69***	-48

SOURCE AND NOTES: See Table 4.1.

[a] "Lower prior earnings" is defined for applicants as earnings of less than $3000 in the year prior to random assignment; for recipients it is zero earnings. "Higher Prior AFDC" means any prior AFDC for applicants and more than two years for recipients. The regression model utilized differs from that employed previously by the introduction of a interaction term for the subgroup combination.

	Prior-Year Earnings		
	None	$1-2,999	$3,000 or More
Had Own AFDC Case			
Never	no	no	no
Two Years or Less	yes	yes	no
More Than Two Years	yes	yes	no

The more dependent subgroup of applicants, therefore, contains returnees with prior-year earnings of less than $3,000. The less dependent subgroup of applicants includes those applicants with no welfare history and those applicant returnees with high prior-year earnings.

A similar split was made for recipients: The less dependent subgroup contains all individuals with any earnings or with prior welfare experience of two years or less. The more dependent subgroup contains those with no earnings and more than two years of prior welfare experience.

Two additional pairs of subgroups were created by moving individuals with a high school diploma from the more dependent groups to the less dependent groups in both the applicant and recipient categories; the more dependent recipients in this categorization are the most dependent subgroups of all. In Table 4.4, the more dependent subgroup of each pair is indicated with a "yes" label and is displayed as the lower row. The Cook County sample was excluded because there was no information on length of prior welfare history.

The results follow closely from those already discussed. With respect to earnings impacts among applicant subgroups, the results for the top pair of applicant subgroups indicate that the less dependent have below average impacts in three out of four programs, and the more dependent have above average impacts. Adding education to the combination of characteristics to increase the dependency contrast indicates, however, that after a certain point on the dependency scale, the impacts become smaller again. The implication is that earnings impacts per enrollee at first tend to increase as dependency increases, but eventually stabilize or begin to decline as the most dependent end of the spectrum is approached.

Among recipients, earnings impacts were somewhat larger for the more dependent half of each pair, but the differences were not as large for applicants. The impacts on AFDC payments were typically not statistically significant, although the pattern for applicants suggests that the impacts may be greater for the more dependent.

II. Further Discussion of Subgroup Differences Across Programs

In comparisons across programs, characteristics other than prior earnings and welfare receipt -- such as education and numbers of children -- usually produced conflicting relationships across program samples. They were, however, sometimes helpful in interpreting interactions between individual characteristics and specific features of each program.

Table 4.5 presents impacts on earnings and welfare payments for applicants and recipients in the five programs under discussion, according to education, numbers and ages of children, marital status, and other characteristics. As before, the impact estimates start at the fourth quarter after random assignment and average all quarters through to the end of the observation period for each program. The few differences between immediate and longer-term impacts that are important are pointed out where relevant.

TABLE 4.5

AFDC APPLICANTS AND RECIPIENTS: UNCONDITIONAL IMPACTS ON EARNINGS AND AFDC PAYMENTS, BY PROGRAM, MINOR SUBGROUP, AND WELFARE STATUS

Subgroup, Welfare Status, and Program	Percent of Sample[a]	Average Earnings Per Quarter Quarters 4 - Last ($)			Average AFDC Payments Per Quarter Quarters 4 - Last ($)		
		Experimental	Control	Difference	Experimental	Control	Difference
High School Diploma							
Applicants							
Yes							
San Diego	61.5	1068	923	+146**	375	420	- 44
Baltimore	44.9	1199	1106	+ 92	337	337	- 0
Virginia	50.8	939	912	+ 27	175	155	+ 21
Arkansas	55.1	557	397	+161**	118	127	- 9
Cook County[b]	46.8	958	957	+ 1	442	457	- 15
No							
San Diego	38.5	609	534	+ 74	532	547	- 15
Baltimore	55.1	829	593	+236***	391	416	- 25
Virginia	49.2	694	559	+136*	206	267	- 61**
Arkansas	44.9	313	270	+ 43	143	190	- 47**
Cook County[b]	53.2	461	532	- 71	495	510	- 16
Recipients							
Yes							
Baltimore	42.1	645	598	+ 46	557	546	+ 10
Virginia	38.8	650	569	+ 80	390	378	+ 12
Arkansas	42.0	180	212	- 32	245	309	- 64*
Cook County[b]	30.9	532	467	+ 66*	663	684	- 20
No							
Baltimore	57.9	347	317	+ 30	679	677	+ 2
Virginia	61.2	363	302	+ 61	426	474	- 48**
Arkansas	58.0	129	73	+ 56	311	369	- 58**
Cook County[b]	69.1	269	230	+ 39	761	771	- 10
Child 12 or Under							
Applicants							
No							
San Diego	22.6	1001	678	+323***	293	341	- 48
Baltimore	27.5	954	942	+ 12	256	244	+ 12
Virginia	22.9	815	636	+180	133	160	- 28
Arkansas	9.6	384	206	+178	83	115	- 32
Yes							
San Diego	77.4	858	803	+ 55	477	506	- 29
Baltimore	72.5	1011	780	+232***	408	432	- 24
Virginia	77.1	821	771	+ 49	208	224	- 17
Arkansas	90.4	455	356	+100*	135	160	- 25*
Recipients							
No							
Baltimore	13.4	271	242	+ 28	438	460	- 22
Virginia	23.7	418	355	+ 63	258	286	- 28
Arkansas	11.8	37	58	- 21	171	251	- 79
Yes							
Baltimore	86.6	504	466	+ 38	657	647	+ 10
Virginia	76.3	492	421	+ 70	459	482	- 23
Arkansas	88.2	165	141	+ 24	298	356	- 58**

(continued)

TABLE 4.5 (continued)

Subgroup, Welfare Status, and Program	Percent of Sample	Average Earnings Per Quarter Quarters 4 - Last ($)			Average AFDC Payments Per Quarter Quarters 4 - Last ($)		
		Experimental	Control	Difference	Experimental	Control	Difference
Number of Own Children							
Applicants							
One							
San Diego	49.7	887	807	+ 80	346	355	- 9
Baltimore	50.4	1033	766	+267***	299	327	- 28
Virginia	49.6	781	696	+ 85	174	185	- 11
Arkansas	42.5	427	391	+ 36	106	129	- 23
More Than One							
San Diego	50.3	895	740	+155**	525	580	- 56*
Baltimore	49.6	966	891	+ 75	434	434	+ 0
Virginia	50.4	856	780	+ 76	207	234	- 27
Arkansas	57.5	467	305	+161**	148	176	- 28
Recipients							
One							
Baltimore	43.1	528	493	+ 35	503	522	- 19
Virginia	42.0	482	426	+ 56	325	341	- 17
Arkansas	34.8	125	173	- 48	192	257	- 65*
More Than One							
Baltimore	56.9	430	392	+ 38	721	696	+ 24
Virginia	58.0	469	391	+ 77	475	504	- 29
Arkansas	65.2	164	108	+ 56	332	390	- 58**
Currently Married							
Applicants							
Yes							
San Diego	46.6	852	750	+102	469	450	+ 19
Baltimore	50.4	951	829	+122	362	374	- 12
Virginia	49.3	820	667	+152*	171	176	- 5
Arkansas	31.9	379	274	+105	117	152	- 35
No							
San Diego	53.4	925	794	+131*	406	482	- 76**
Baltimore	49.6	1046	824	+222***	370	387	- 16
Virginia	50.7	819	810	+ 9	210	243	- 33
Arkansas	68.1	481	373	+108*	136	158	- 22
Recipients							
Yes							
Baltimore	34.3	436	422	+ 14	614	617	- 3
Virginia	38.3	464	427	+ 37	433	422	+ 11
Arkansas	21.2	173	242	- 69	311	293	+ 18
No							
Baltimore	65.7	490	442	+ 48	634	624	+ 10
Virginia	61.7	481	394	+ 87*	398	442	- 44**
Arkansas	78.8	145	103	+ 43	274	356	- 82***

(continued)

TABLE 4.5 (continued)

Subgroup, Welfare Status, and Program	Percent of Sample	Average Earnings Per Quarter Quarters 4 - Last ($)			Average AFDC Payments Per Quarter Quarters 4 - Last ($)		
		Experimental	Control	Difference	Experimental	Control	Difference
Ever Married							
Applicants							
Yes							
San Diego	84.1	908	806	+102*	421	445	- 24
Baltimore	69.9	1003	821	+182**	346	363	- 17
Virginia	74.2	830	705	+125*	170	176	- 6
Arkansas	56.9	424	310	+114*	112	123	- 11
No							
San Diego	15.9	805	608	+196	511	591	- 81
Baltimore	30.1	980	832	+148	413	421	- 8
Virginia	25.8	789	841	- 52	249	308	- 59*
Arkansas	43.1	481	383	+ 98	153	199	- 45**
Recipients							
Yes							
Baltimore	49.1	458	425	+ 33	621	624	- 2
Virginia	65.3	462	376	+ 86*	400	402	- 2
Arkansas	41.8	159	158	+ 1	329	328	+ 1
No							
Baltimore	50.9	487	446	+ 40	633	620	+ 13
Virginia	34.7	497	459	+ 38	432	495	- 63**
Arkansas	58.2	143	112	+ 32	249	353	-104***
Age							
Applicants							
30 or Over							
San Diego	65.6	966	776	+189***	411	464	- 54*
Baltimore	65.4	1076	927	+150*	349	347	+ 2
Virginia	64.0	858	741	+117*	183	183	+ 1
Arkansas	38.5	392	330	+ 62	121	138	- 17
Cook County[b]	62.1	763	860	- 98*	451	453	- 2
Less Than 30							
San Diego	34.4	747	770	- 22	484	477	+ 7
Baltimore	34.6	846	633	+213**	398	442	- 44
Virginia	36.0	750	737	+ 13	204	260	- 56**
Arkansas	61.5	484	348	+135**	135	167	- 32*
Cook County[b]	37.9	581	519	+ 62	500	539	- 39
Recipients							
30 or Over							
Baltimore	42.7	398	394	+ 4	636	632	+ 3
Virginia	65.9	432	401	+ 31	411	426	- 15
Arkansas	38.1	139	68	+ 71	324	356	- 32
Cook County[b]	43.1	326	278	+ 48	765	792	- 27*
Less Than 30							
Baltimore	57.3	527	467	+ 61	621	614	+ 7
Virginia	34.1	557	419	+138**	412	453	- 41
Arkansas	61.9	161	173	- 13	261	338	- 77***
Cook County[b]	56.9	368	325	+ 43	705	707	- 2

TABLE 4.5 (continued)

Subgroup, Welfare Status, and Program	Percent of Sample	Average Earnings Per Quarter Quarters 4 - Last ($)			Average AFDC Payments Per Quarter Quarters 4 - Last ($)		
		Experimental	Control	Difference	Experimental	Control	Difference
Recent UI Benefits							
Applicants							
Some							
San Diego	14.1	1270	1304	- 34	390	471	- 80
None							
San Diego	85.9	828	684	+144***	443	467	- 24
Labor Market							
Applicants							
Urban							
Virginia	78.7	849	769	+ 80	195	213	- 18
Arkansas	64.8	403	302	+100	99	126	- 27
Rural							
Virginia	21.3	707	625	+ 82	174	198	- 24
Arkansas	35.2	533	413	+120	187	211	- 24
Recipients							
Urban							
Virginia	78.8	521	421	+100**	411	450	- 39**
Arkansas	56.5	211	120	+ 91*	250	325	- 75***
Rural							
Virginia	21.2	299	346	- 47	413	383	+ 30
Arkansas	43.5	76	152	- 76	326	366	- 41

SOURCE AND NOTES: See Table 4.1.

[a] Percent of applicants and percent of recipients.

[b] The definitions of "applicant" and "recipient" for Cook County differ from the other programs. See text for discussion.

[c] For Baltimore, Virginia, and Arkansas the category "black" includes a small number of individuals in other non-white groups. In San Diego and Cook County, "white" includes a small number of non-black, non-hispanic, non-white persons.

TABLE 4.5 (continued)

Subgroup, Welfare Status, and Program	Percent of Sample	Average Earnings Per Quarter Quarters 4 - Last ($)			Average AFDC Payments Per Quarter Quarters 4 - Last ($)		
		Experimental	Control	Difference	Experimental	Control	Difference
Ethnicity							
Applicants							
White							
San Diego[c]	61.5	949	821	+128*	357	369	- 12
Baltimore	33.8	922	767	+155	309	318	- 9
Virginia	41.8	801	663	+138	128	149	- 21
Arkansas	16.7	403	259	+144	72	82	- 11
Cook County[b,c]	21.8	679	709	- 30	353	413	- 61*
Black							
San Diego	20.7	895	589	+306***	532	678	-146***
Baltimore[c]	66.2	1035	855	+180**	395	412	- 17
Virginia[c]	58.2	832	792	+ 40	235	253	- 18
Arkansas[c]	83.3	458	358	+100*	141	170	- 29*
Cook County[b]	65.5	741	761	- 20	502	517	- 15
Hispanic							
San Diego	17.8	693	843	-150	593	556	+ 38
Cook County[b]	12.6	458	608	-150	510	450	+ 60
Recipients							
White							
Baltimore	25.1	420	438	- 18	579	575	+ 4
Virginia	26.8	490	398	+ 92	303	305	- 2
Arkansas	8.5	337	55	+282**	207	228	- 21
Cook County[b]	14.3	517	322	+195***	572	604	- 31
Black							
Baltimore[c]	74.9	490	436	+ 55	644	638	+ 6
Virginia[c]	73.2	468	408	+ 60	451	483	- 32*
Arkansas[c]	91.5	132	136	- 4	290	354	- 64***
Cook County[b]	75.3	322	307	+ 15	761	772	- 11
Hispanic							
Cook County[b]	10.4	324	265	+ 59	730	734	- 4
Sex							
Applicants							
Male							
Cook County[b]	10.2	867	931	- 64	432	369	+ 63
Female							
Cook County[b]	89.8	673	708	- 35	474	499	- 25
Recipients							
Male							
Cook County[b]	13.3	451	288	+163***	704	746	- 42
Female							
Cook County[b]	86.7	334	307	+ 27	735	743	- 8

(continued)

Information not shown in Table 4.5 is also used in the discussion where relevant in explaining the impact results.

A. San Diego

The welfare employment program in San Diego, as Chapter 2 indicated, was distinctive in two important respects. First, the program served only welfare applicants. Second, all enrollees had the same short-term sequence of program activities -- job search followed by work experience for those who did not find a job. Participation rates were high for all subgroups.

The San Diego program clearly had greater impacts on the less job-ready and more welfare-dependent applicants. Those with zero prior earnings had by far the largest earnings impacts; the welfare savings were spread evenly over those with zero prior earnings and those with earnings that were positive but under $3,000. Similarly, applicants with a welfare history had most of the earnings gains and welfare savings, although both impacts were somewhat greater for the group with a welfare history of two years or less than for those with more than two years.

Some characteristics associated with dependency other than prior earnings and welfare history appear to be positively related to the program's impacts in San Diego. The results for subgroups presented in Table 4.5 suggest, for example, that ethnicity and the number of children in a household were factors associated with greater dependency for this sample, as reflected in the higher welfare payments made to control group members who were non-white and had more than one child.

Some of the other subgroup comparisons, however, were not consistent with a greater effect on the more dependent -- notably the greater impacts for applicants who had a high school diploma or GED, a factor not usually related to long-term dependency. This result may stem from the reliance on job search in San Diego. More education may have increased the probability of success in a program that (unlike Baltimore) did not offer remedial education.

B. Baltimore

The Options program in Baltimore was very different from the San Diego initiative. Newly-mandatory AFDC recipients (with their youngest child just entering school) were enrolled as well as mandatory applicants. In addition, there was a wider range of services -- from independent job search to education and training -- and the services could vary according to the registrants' needs and preferences. Finally, enrollment during the period of the study was limited to 1,000 slots per year to ensure that the resources the planners deemed adequate to provide this range of services were available. As it turned out, the cost per experimental was estimated to be substantially higher in Baltimore than elsewhere, approximately $1,050 (in 1987 dollars). Participant choice was constrained by slot availability and guided by staff. Because the least job-ready generally participated at higher levels in the more intensive services -- education and training -- than did other subgroups, the subgroup impacts may have been influenced by the different services participants received, as well as by participants' own characteristics.

The Baltimore results on which this study is based make use of an extra year of follow-up data that was not available for the MDRC final report on the Options program.[3] Analysis of those additional data showed somewhat larger earnings impacts than even the initial favorable short-term estimates presented in the program.[4] Earnings impacts increased after the first follow-up year. In fact, most of the program effects on earnings accrued after the first year. Welfare savings, which were small in the initial follow-up period, did not increase.

For Baltimore applicants, results for the major subgroups fit the pattern already discussed for San Diego. The least dependent groups improved the least. Of additional interest, there was a relatively wide spread between earnings gains for applicants and for recipients. As a group (see Table 4.1 earlier in the chapter), applicants in the program earned $172 more per quarter than controls, a statistically significant increase of 21 percent that is comparable to the change for applicants in San Diego. However, recipients -- the more welfare-dependent -- earned only $37 more. The difference in earnings impacts between applicants and recipients is statistically significant.

These findings are especially important because Baltimore recipients had high participation rates and, as noted in Chapter 2, gross costs of serving a recipient were 25 percent higher than for serving an applicant. Recipients received a somewhat larger share of the more expensive services, and the follow-up period was long enough to capture the post-program effects of education and training. Most of the lack of impact for recipients was due to the lack of improvement for the most- and least-dependent subgroups among the recipient category. Recipients with prior employment showed negative earnings impacts, and recipients with more than two years on the rolls combined with no recent work experience also showed below-average improvement. Together, these subgroups make up most of the recipient sample.

With respect to impacts for subgroups defined by education and demographic characteristics, the Baltimore results were opposite to those in San Diego. Applicants with less education had larger than average impacts, perhaps reflecting the remedial education services offered by the Baltimore Options program. Younger women, women with younger children and women with fewer children also experienced somewhat larger-than-average gains among applicants. Such contrasting results across programs make it difficult to predict the effect of these characteristics under a variety of program settings.

C. Virginia

Virginia extended program participation requirements to the whole WIN-mandatory caseload of recipients as well as mandatory AFDC applicants. It also served rural as well as urban areas, and counties had considerable independence in implementing the program.[5] Resource constraints were important, however. The counties relied on job search assistance as their principal component and on independent job search as the most widely-used kind of job search. Community providers, such as schools and JTPA training programs, which received no program funding, were utilized on a referral basis for education and training activities. Since controls obtained these education and training services on their own about as much as experimentals, however, it is not clear that this component contributed much to the estimated Virginia impacts.[6]

Impacts on employment and earnings in Virginia are interesting for the several anomalies they present. First, although short-term earnings impacts for applicants were larger than for recipients, this differential largely disappeared in the long-term follow-up period analyzed in this study. Second, applicants without recent earnings, who elsewhere had average impacts or above-average impacts, experienced virtually a zero longer-term effect in Virginia. Finally, the most job-ready subgroup, applicants with $3,000 or more in prior-year earnings, obtained the largest earnings impacts.

To examine this constellation of results further, Figure 4.1 gives the quarter-by-quarter employment impacts for the three prior-earnings applicant subgroups. As shown, the experimental-control differential for the no-prior-earnings subgroup was relatively wide during the first six follow-up quarters, but then began to decline as controls "caught up." Thus, the short-term nature of the employment impacts, not their total absence, was responsible for the poor showing of this subgroup. It was also the primary reason why overall effects for applicants in Virginia are not greater than the effects for recipients there. As shown, the middle subgroup also experienced impact decay, although the top subgroup did not.

It is not obvious why the Virginia program should have had its largest employment and earnings impacts on a top employability subgroup. But some consequences of that result are worth noting. Table 4.6 provides additional impact information for the subgroup with at least $3,000 in prior-year earnings, by showing a breakdown of the subgroup's employment status during the longer-term follow-up by welfare status at the end of the follow-up period. This breakdown covers all four possible combinations of work and welfare receipt: did not work and remained on welfare, did not work and left welfare, worked and left welfare, and worked but remained on welfare. The table shows an 8.4 percentage point increase in the number of individuals in this subgroup who worked after the third follow-up quarter and were off welfare at the last quarter (the eleventh). Corresponding to this

FIGURE 4.1

VIRGINIA APPLICANTS: EMPLOYMENT IMPACTS BY PRIOR EARNINGS SUBGROUP

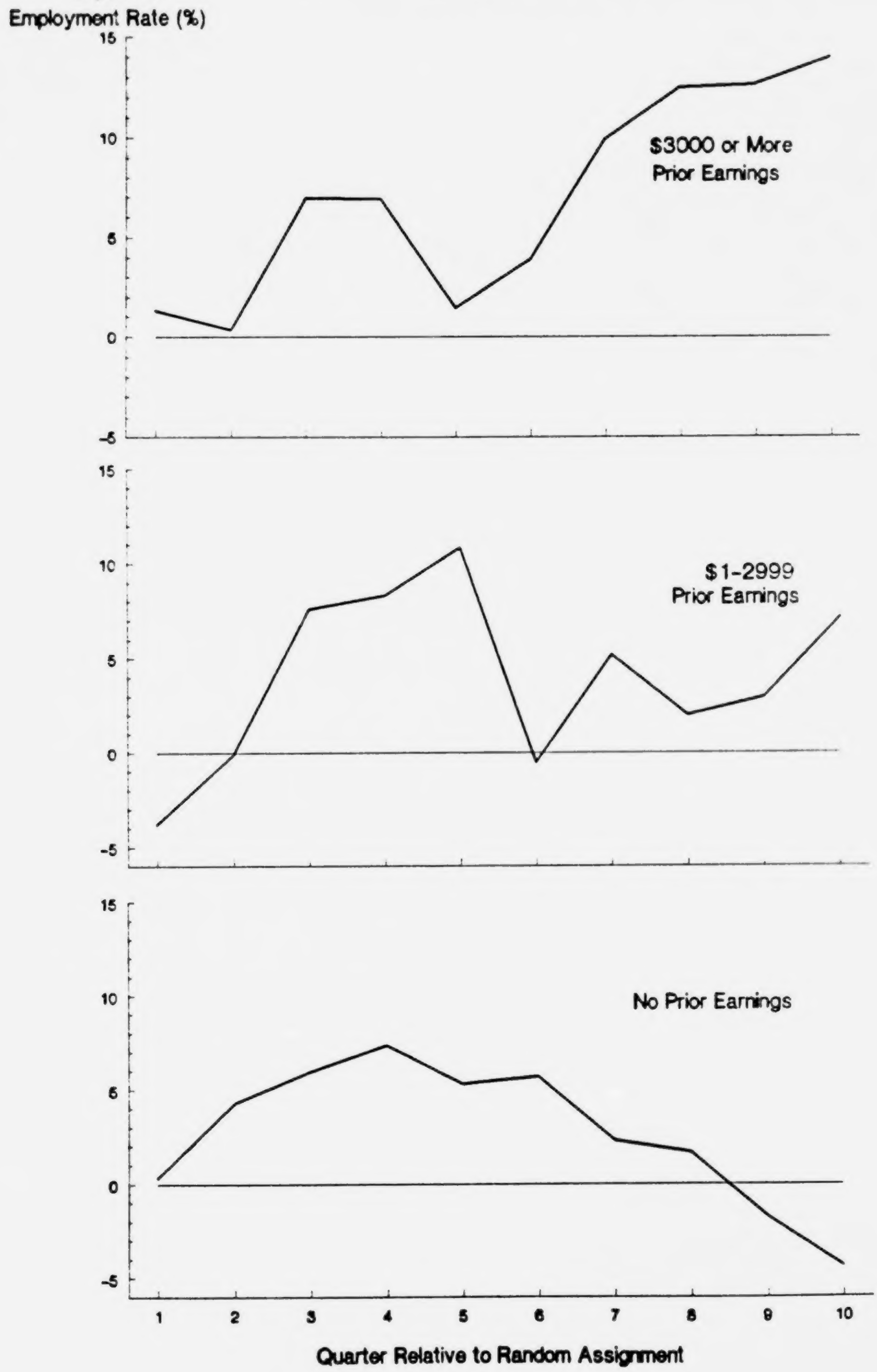


TABLE 4.6

VIRGINIA

AFDC APPLICANTS HAVING EARNINGS OF $3,000 OR MORE IN THE YEAR PRIOR TO RANDOM ASSIGNMENT: EMPLOYMENT STATUS IN QUARTERS 4 - LAST BY WELFARE STATUS IN QUARTER 11

Employment and Welfare Status (%)	Experimentals	Controls	Difference
Had No Earnings, Received No AFDC Payments	6.1	13.5	- 7.5*
Had Some Earnings, Received No AFDC Payments	75.7	67.2	+ 8.4
Had No Earnings, Received Some AFDC Payments	4.9	5.6	- 0.7
Had Some Earnings, Received Some AFDC Payments	13.4	13.7	- 0.3
Total	100.0	100.0	0.0
Sample Size	244	127	

SOURCE: MDRC calculations from the Commonwealth of Virginia Unemployment Insurance earnings records, welfare records from the Virginia Automated Client Information System, and Fairfax County AFDC case files.

NOTES: These data are regression-adjusted using ordinary least squares, controlling for pre-enrollment characteristics of sample members. There may be some discrepancies in calculating sums and differences due to rounding.

A two-tailed t-test was applied to differences between Experimental and Control groups. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; *** = 1 percent. The distributed differences are not, however, strictly independent.

116

increase, however, was a 7.5 percentage point decrease in the number of individuals who went off welfare without working.

Thus, the effect of the program for this subgroup was to increase employment for those who would have been off welfare relatively quickly anyway. Only a minority -- under 20 percent -- of this subgroup would have remained on welfare even without special intervention, and the program had little impact on them. Most of the program impact evidently "spilled over" onto those who would only have been on AFDC for a short period anyway. As a consequence, welfare dependency was not affected.

The point of this discussion is that even where programs have above-average earnings impacts on the most job-ready enrollees, the result may not fulfill all program objectives. On the surface, the findings for Virginia appear to provide support for targeting the most job ready. In fact, they demonstrate another argument against making these individuals the exclusive focus of resources and attention. That is, earnings gains to them appear to have little effect on welfare dependency.

Results for the subgroups defined by individual characteristics in Virginia (see Table 4.5 above) show that, although the differences in impacts among subgroups were typically similar in size to those in the other states, they were often not in the same direction. Thus, as already indicated, few cross-state generalizations can be made about impact differences according to characteristics other than prior employment and welfare history. It is worth noting that the effects for applicants without a diploma or GED were larger than for those with one. This may be connected with the program's use of referral to education and training providers. Owing to the high incidence of similar activities among controls in Virginia, however, this can be only speculation.

D. Arkansas

The Arkansas program was also heavily constrained by resources. The program formally offered job search and work experience but, in fact, rates of participation in work experience were low. Program cost per experimental was among the lowest of all programs in MDRC's Work/Welfare Demonstration.

Arkansas provides an opportunity to examine the effects of a welfare employment program in a low-grant state. At the start of the research, the maximum welfare benefit for a mother with two children was $140 per month, a benefit low enough to make even a small amount of earnings disqualify a family for AFDC. A modest increase in employment might therefore be expected to have a relatively large effect on welfare receipt in this state. As it turned out, the welfare reductions were the largest of the programs evaluated, whether the reductions are measured as percent on welfare, absolute dollar payments, or payments as a percentage of control group payments.

The low benefit levels also determine the nature of the sample. Individuals relying on welfare were likely to have only very limited opportunities for income elsewhere. The sample in Arkansas may therefore be more disadvantaged than in other states in the study. Some evidence of this is the relatively low follow-up employment rates for the Arkansas sample compared to the other programs.

The Arkansas sample is of special interest for one other reason. The state obtained federal waivers to extend mandatory program coverage to AFDC's whose youngest child was three to five years old. As it turned out, about half the sample fell into this category. The data therefore afford an opportunity to see whether the presence of a pre-school child affects how much a woman is likely to benefit from the program. The opportunity is limited, however, because the research sample for Arkansas has only 1,127 members, constraining the potential for sample subdivision.

The earnings impacts for major employability and dependence subgroups in Arkansas fit the general pattern. Earnings impacts were small for applicants who never had their own AFDC case before and for applicants with year-prior earnings of $3,000 or more. Applicant returnees had the largest earnings impacts. The recipient subgroups had relatively weak earnings impacts.

Applicants without a child under 12 and applicants who were white had earnings increases somewhat above average, as in Virginia. Both these characteristics are associated with lower welfare dependency. Applicants with a high school diploma also had relatively large earnings gains. When these three characteristics were combined in a regression framework with information on other applicant characteristics, the maximum earnings impact occurred with the least dependent, with impacts decreasing gradually as dependency increased. This pattern was also similar to Virginia's in that the earnings gains for the least dependent were not accompanied by welfare reductions for them.

In contrast to the pattern for earnings gains, longer-term welfare reductions were largest among recipients (Table 4.1). They were large among the most dependent recipient subgroups and subgroup combinations (Table 4.4), and exceeded the earnings increases that came to these subgroups during the same period of follow-up. Additional analysis confirmed that the welfare savings for recipients were not the result of any sustained increase in employment. Employment gains for recipients peaked in quarter 3, and decayed after that. Most of this decay was associated with the absence of a program impact on the length of employment.[7] This raises questions about the overall financial effects on enrollees. For recipients, three quarters or more of the impact on AFDC payments came through program effects other than a sustained increase in UI earnings.[8] The welfare reduction was therefore not offset by an increase in the enrollees' own earnings. Thus, although sample members near the least dependent end of the continuum showed a net increase in the total of their own earnings plus AFDC payments, those who were the most dependent showed a net loss from these two sources taken together.

This mismatch between earnings and welfare impacts among the more dependent was not found generally in the study samples elsewhere, but it nevertheless raises an important monitoring and targeting issue. Performance standards are usually thought of as tools to maximize program effects on earnings _or_ on welfare receipt; the pattern in Arkansas suggests that standards should also be tested to determine whether they maximize effects on _both_ outcomes at once. This issue is discussed further in connection with performance standards in the next chapter.

Some comment on impact differences by demographic characteristics is warranted. First, as shown in Table 4.5 above, earnings gains for the more rural of the two study locations were lower than for Little Rock, essentially because the effects for rural recipients were negative. This is consistent with findings for the rural counties of Virginia, where recipients also had negative earnings effects (see Table 4.5 above). Evidence from an

MDRC evaluation of a work experience program in West Virginia is also relevant in this connection.[9] This state is largely rural and during the research experienced some of the nation's highest unemployment rates. The impact study covered primarily AFDC recipients, and found virtually no effects on employment or earnings. The weight of this evidence is not by any means conclusive, but it suggests that these kinds of interventions may not have strong impacts among longer-term AFDC recipients in rural areas.

Second, the welfare savings for women who were not married, particularly those who were never married, are considerably larger than for married or separated women. Unmarried and never-married women normally remain on welfare longer than married women, but the evidence from this study suggests that there exists a potential for reducing this greater dependency. As noted earlier, this finding is one of the few with some consistency across programs for the subgroups defined by demographic characteristics other than prior earnings or prior welfare. Unmarried women obtained virtually all the realized welfare savings in San Diego, Virginia, and Arkansas; of these, never-married women obtained over 70 percent.[10] There were no welfare savings in Baltimore. The Cook County sample is excluded from these calculations, since marital status was not available for the impact sample there.

Impact estimates for mandatory mothers with pre-school children are shown in Table 4.7. Average employment rates for controls indicate that the presence of a young child did not reduce employment in the absence of the program. Nor was it associated with lower impacts on employment or earnings. In fact, for the full sample, impacts on employment and earnings were larger for women with pre-school children, although this extra effect

TABLE 4.7

ARKANSAS

AFDC APPLICANTS AND RECIPIENTS: UNCONDITIONAL IMPACTS ON EMPLOYMENT, EARNINGS, AFDC INCIDENCE AND PAYMENTS, BY WELFARE STATUS AND CHILD LESS THAN SIX SUBGROUP

Subgroup and Welfare Status	Percent of Sample[a]	Percent Employed Quarterly Quarters 4 - Last (%)			Average Earnings Per Quarter Quarters 4 - Last ($)		
		Experimental	Control	Difference	Experimental	Control	Difference
Any Child Less Than 6							
Full Sample							
No	46.0	21.9	18.3	+3.6	309	254	+55
Yes	54.0	24.9	18.1	+6.8***	342	259	+83*
Applicants							
No	51.2	27.1	22.6	+4.5	404	339	+65
Yes	48.8	34.6	24.1	+10.4***	496	345	+152**
Recipients							
No	38.3	11.5	9.6	+1.8	125	84	+41
Yes	61.7	14.1	11.1	+3.0	166	161	+ 5

Subgroup and Welfare Status	Percent of Sample[a]	Percent Receiving AFDC Monthly Quarters 4 - Last (%)			Average AFDC Payments Per Quarter Quarters 4 - Last ($)		
		Experimental	Control	Difference	Experimental	Control	Difference
Any Child Less Than 6							
Full Sample							
No	46.0	32.6	40.3	-7.7**	168	207	-39**
Yes	54.0	39.5	46.0	-6.4**	212	253	-41**
Applicants							
No	51.2	23.3	28.4	-5.0	119	145	-26
Yes	48.8	26.5	30.9	-4.4	141	167	-26
Recipients							
No	38.3	50.5	63.2	-12.6**	263	327	-64*
Yes	61.7	54.5	63.8	- 9.4**	295	354	-58**

SOURCE AND NOTES: See Table 4.1.

[a] Percent of full sample, applicants, or recipients.

was not statistically significant.[11] About 64 percent of the total program effect on earnings came from working with this subgroup. The differential appears to vary across applicant and recipient categories, but this variability may stem from the small number of sample points in these subdivisions of the Arkansas sample. Welfare savings were almost identical for women with and without a pre-school child. Once again, it is worth emphasizing that caution should be exercised in generalizing these findings owing to the small sample size and special program environment.

E. Cook County

In contrast to programs like San Diego and Baltimore, Cook County sought to involve the full WIN-mandatory population in some activity. In the absence of substantial supplementary funding, this decision necessarily meant that Cook County's limited resources were spread over one of the largest caseloads in the nation -- more than 50,000 WIN-mandatory AFDC recipients at any one time. The resulting program expenditure of $150 (in 1987 dollars) per experimental was lower than any other state evaluated -- slightly under the cost for Arkansas and less than one-sixth the cost for Baltimore.

The effort to reach the full WIN-mandatory caseload with limited resources, combined with a belief that the best way to foster independence was to make program participants take responsibility for their own activities, shaped the character of the program. First, the program stipulated that individual clients assume the primary responsibility for finding jobs on their own. Independent Job Search was selected as the primary activity, although individuals who completed IJS without finding a job could then be required to participate in work experience for three months. Second, program staff, whose caseloads averaged about 300, were evaluated largely on the basis of the AFDC grant reductions they achieved. They therefore tended to concentrate more on administrative and monitoring functions than on direct services to clients; sanctioning for failure to satisfy program requirements, for example, was more automatic than in other programs studied by MDRC. In sum, the balance between enforcing obligations and providing services weighed more heavily towards the former in Cook County than elsewhere.

As noted, data distinguishing between applicants and recipients were lacking in Cook County. As a substitute, individuals who received no AFDC payments during the three months preceding random assignment are considered applicants; the remainder (about two-third of the sample) are considered recipients.[12] As also noted, this limitation alone would probably not reduce comparability with the other program subgroups substantially. The important difference is that in Cook County, only approved applicants and recipients were enrolled. For this reason, applicants in Cook County had higher rates of continuing welfare receipt than applicants in other program samples, falling about midway between the other applicant and recipient samples.

Other adjustments in subgroup definitions were also necessary for Cook County. As stated earlier, neither background demographic nor welfare history data were available. AFDC grant payments were available for a ten-month period preceding random assignment. These were used to subdivide the large category of recipients with no prior-year earnings. The first subdivision was between those who did not receive welfare payments in all ten prior months and those who did. This latter groups was further divided into three groups depending on payment amount.

The sample was thereby broken down into eight mutually exclusive categories, shown in Table 4.8 in descending order of welfare dependency. The table gives earnings and welfare impacts for these subgroups. As stated in the final evaluation report for the Cook County program, there were no statistically significant employment and earnings impacts for the full sample, although there were small welfare savings.[13] The table indicates that increases in employment and earnings were obtained for one subgroup in the middle range of dependency, namely, recipients with prior employment.[14] This subgroup accounted for 13 percent of the sample, but for virtually all the earnings impact and 35 to 40 percent of the aggregate welfare savings. Quarterly earnings gains for them were nearly $150 larger than their AFDC payments reductions. Neither the subgroups above them on the dependency scale, nor the large body of individuals in the lower half of the dependency scale gained much from the program.

Of the demographic characteristics, only age, sex and highest grade obtained at time of welfare application were available (Table 4.5). Consistent with the earnings gains observed for recipients with prior earnings, Table 4.5 shows statistically significant earnings impacts for recipients with a high school diploma, and also shows relatively large earnings gains for the small subgroups of recipient males and recipient whites. These were all less-dependent subgroups within the recipient category.

III. Summary for the Major Subgroups and Combinations

By way of summary, Table 4.9 displays the earnings and AFDC payment

TABLE 4.8

COOK COUNTY

AFDC APPLICANTS AND RECIPIENTS: UNCONDITIONAL IMPACTS ON EMPLOYMENT, EARNINGS, AFDC INCIDENCE AND PAYMENTS, BY WELFARE STATUS, PRIOR EARNINGS, AND PRIOR WELFARE PAYMENTS

Subgroup, Welfare Status, and Program	Percent of Sample[a]	Percent Employed Quarterly Quarters 4 - Last (%)			Average Earnings Per Quarter Quarters 4 - Last ($)		
		Experimental	Control	Difference	Experimental	Control	Difference
Applicants[b]							
Prior Earnings							
$3000 or More	11.1	53.9	52.2	+ 1.7	1373	1336	+ 37
$1-2999	6.9	33.5	38.4	- 4.8	555	722	-167
None	15.7	15.0	16.5	- 1.5	268	301	- 33
Recipients[b]							
Some Prior Earnings	13.1	42.2	38.0	+ 4.2**	865	685	+179***
No Prior Earnings							
1-9 Months Prior AFDC	6.8	15.7	14.3	+ 1.4	252	222	+ 30
10 Months Prior AFDC							
Average Payment							
Less Than $3000	18.8	15.8	16.5	- 0.7	242	284	- 42
$3000-3499	12.5	13.4	11.6	+ 1.8	219	196	+ 24
$3500 or More	15.1	12.8	9.3	+ 3.5*	188	125	+ 63

Subgroup, Welfare Status, and Program	Percent of Sample[a]	Percent Receiving AFDC Monthly Quarters 4 - Last (%)			Average AFDC Payments Per Quarter Quarters 4 - Last ($)		
		Experimental	Control	Difference	Experimental	Control	Difference
Applicants[b]							
Prior Earnings							
$3000 or More	11.1	46.4	47.0	- 0.6	406	403	+ 3
$1-2999	6.9	56.7	59.3	- 2.6	460	493	- 33
None	15.7	59.4	62.8	- 3.4	519	540	- 21
Recipients[b]							
Some Prior Earnings	13.1	62.1	67.0	- 4.8**	567	607	- 40*
No Prior Earnings							
1-9 Months Prior AFDC	6.8	66.8	68.7	- 1.9	613	620	- 7
10 Months Prior AFDC							
Average Payment							
Less than $3000	18.8	76.1	77.0	- 0.9	601	596	+ 5
$3000-3499	12.5	83.1	82.6	+ 0.5	815	820	- 6
$3500 or More	15.1	85.5	86.1	- 0.7	1020	1041	- 21

SOURCE AND NOTES: See Table 4.1.

[a] Percent of full sample.

[b] The definitions of "applicant" and "recipient" for Cook County differ from the other programs. See text for discussion.

TABLE 4.9

AFDC APPLICANTS AND RECIPIENTS: SUMMARY OF IMPACTS FOR MAJOR SUBGROUPS AND COMBINATIONS

	Quarterly Earnings Impact, Quarters 4 - Last ($)				
Subgroup[a]	San Diego	Baltimore	Virginia	Arkansas	Cook County[b]
First Tier					
Applicants With No Prior AFDC	+37	+121	-13	+26	---
Applicants With $3000 or More Prior Earnings	-39	+18	+307***	+12	+37
Second Tier					
Applicant Returnees	+158**	+188***	+114*	+211***	---
Applicants With Less Than $3000 Prior Earnings	+181***	+244***	-16	+121**	-75
Applicant Returnees With Less Than $3000 Prior Earnings	+151**	+253***	+20	+202**	---
Third Tier					
All Recipients	---	+37	+69*	+19	+46**
Recipients With More Than Two Years on AFDC	---	-0	+110**	+14	---
Recipients With No Prior Earnings	---	+104**	+70	+29	+12
Recipients With No Prior Earnings and More Than Two Years on AFDC	---	+88	+94*	+28	---
All AFDC					
Quarterly Earnings Impact	+118**	+96***	+72**	+70**	+19
Average Control-Group Earnings	773	634	541	257	451

(continued)

TABLE 4.9 (continued)

Subgroup[a]	Quarterly AFDC Payment Impact, Quarters 4 - Last ($)				
	San Diego	Baltimore	Virginia	Arkansas	Cook County[b]
First Tier					
Applicants With No Prior AFDC	-5	-9	-28	-31	---
Applicants With $3000 or More Prior Earnings	+3	+7	+8	+1	+3
Second Tier					
Applicant Returnees	-47	-15	-16	-19	---
Applicants With Less Than $3000 Prior Earnings	-48*	-23	-31	-30*	-25
Applicant Returnees With Less Than $3000 Prior Earnings	-63*	-19	-29	-22	---
Third Tier					
All Recipients	---	+5	-24	-60***	-13
Recipients With More Than Two Years On AFDC	---	+19	-48**	-44*	---
Recipients With No Prior Earnings	---	+1	-26	-63***	-6
Recipients With No Prior Earnings and More Than Two Years on AFDC	---	-1	-48**	-48*	---
All AFDC					
Quarterly AFDC Payment Impact	-33	-5	-23*	-40***	-13
Average Control-Group AFDC Payments	469	501	345	232	646

SOURCE AND NOTES: See Table 4.1.

[a] Subgroups within each tier overlap, constituting alternative ways of grouping individuals. There is also overlap between first and second tiers.

[b] The definitions of "applicant" and "recipient" for Cook County differ from the other programs. See text for discussion.

impact estimates for several of the major subgroups and combinations discussed. As can be seen, these subgroups are not mutually exclusive. But they are useful because they represent several ways that program planners could distinguish among program eligibles using information about prior work and welfare. The table is organized in three roughly increasing tiers of dependency. Several conclusions may be drawn.

First, subgroups in the top tier (the least dependent subgroups) do not generally exhibit above-average employment and earnings impacts; indeed, they usually exhibit low impacts, if any. These subgroups constitute anywhere from 10 to over 50 percent of applicants. Any program design that focuses the bulk of resources on these subgroups, passing over subgroups further down, is unlikely to be maximizing program impact.

The second tier (the middle dependency tier) consists of applicants either with some welfare history or with a weak work record. Earnings impacts for these subgroups were more likely to be above average, although this was not true for all subgroup samples in this tier. Across samples, applicant returnees evidenced the most consistent gains of any subgroup in this study. In the four programs where welfare history was available they accounted for 65 to 70 percent of the average earnings impact for all groups, although it is unlikely that this subgroup obtained impacts in the fifth program (Cook County).[15] These returnees were individuals who had shown some dependence on AFDC in the past, but who had not received it continuously. This may derive from a greater capacity for independence, with returns to welfare occasioned by situational difficulties or loss of a job -- problems that can often be alleviated by the kinds of work/welfare programs included in the study.

The third tier (the most dependent subgroups) comprises several recipient subgroups. These groups together make up a large share of program eligibles -- 40, 50, and 60 percent of the samples, respectively, in Arkansas, Baltimore, and Virginia, and two-thirds of the sample in Cook County. Employment and earnings impacts were found for some recipient subgroups, but the pattern was not consistent. Lack of impacts was not due to lack of program participation. On the contrary, participation rates were as high as or higher for recipients than for applicants.

This third tier does, however, cover a rather wide range of dependency. The lower part of that range is composed of recipients without recent employment, with more than two years previous welfare receipt, or both. For the four recipient samples, only occasionally did any of these subgroups attain above-average earnings impacts, although some were positive and statistically significant. Therefore, exclusive targeting on the most dependent may be a risky strategy for a program whose major objective is to maximize earnings impacts, at least for the kinds of programs under study.

Welfare savings, because they were smaller than earnings increases, are more difficult to contrast across subgroups. The least dependent groups showed no welfare impacts. For the other two tiers, no strong evidence emerged that favored preferential targeting for any subgroup, although the more dependent subgroups of the samples tended to show larger welfare effects than the other subgroups in some cases.

No subgroup emerges clearly and consistently as the most promising group on which to focus exclusive targeting efforts when sufficient resources are available to serve the bulk of program eligibles. On the other hand, subgroups in the second tier show promise for priority attention when earnings gains are the goal and scarce resources are a constraining factor. Applicant returnees evidenced the most consistent earnings responses, with gains in all four of the programs that produced overall earnings impacts. Applicants in the lower prior earnings brackets and returnees with lower prior earnings are overlapping categories which had higher than average earnings impacts in three of five programs. Program operators who can serve only a portion of the WIN-mandatory caseload may consider one of these subgroups a suitable starting point. The results indicate that exclusive attention to the top tier is clearly a poor strategy if an increase in earnings or a reduction in dependency are program objectives.

The evidence further suggests caution in targeting very narrowly. Applicants as a whole obtained benefits in four of five program samples; progressively narrower targeting within the applicant group does not necessarily yield progressively larger impacts. Operators should also be sensitive to the possibility that the presence of some job ready registrants may itself help program effectiveness by providing encouragement to the less job ready and to program staff.

Interpretation of the findings for recipients is problematic. The findings indicate that restricting program enrollment only to recipients would be a risky strategy for programs whose objective was to maximize earnings impacts. However, a substantial share of the estimated welfare savings were found among recipients. It remains an open question whether shifting resources from the least dependent in order to provide more expensive services to the other subgroups would have increased earnings impacts on those subgroups.

The combination of impact differences across dependency subgroups and according to demographic characteristics suggests a possible threshold effect for earnings. That is, for cases whose multiple disadvantages combine to make them more dependent than some threshold level, the typically low-cost services provided by the programs in this study may begin to lose their effectiveness in raising earnings. The data in support of the threshold effect hypothesis are only suggestive -- other characteristics are associated with dependency, and some dependent subgroups did relatively well in some programs.

Nevertheless, it is at least plausible that there is a substantial group for which the relatively low-cost interventions included in this study lose some of their effectiveness in improving earnings outcomes. This group may include recipients with more than two years on AFDC, with no recent earnings, and without a high school diploma. This combination of characteristics was associated with below-average earnings gains not only in Baltimore, but also in Arkansas and, to a lesser degree, in Virginia. San Diego is not comparable because the evaluation did not include recipients. No information about length of prior welfare was available in Cook County, but such a threshold effect is suggested by the finding that recipients without prior employment had no earnings impacts.

CHAPTER 5

MEASURES OF PROGRAM PERFORMANCES

Our analysis of welfare program performance for different subgroups of the population defines performance in terms of program impacts on employment and welfare of the people it serves. But direct estimates of impacts cannot be obtained cheaply or quickly enough to be used in the management of most programs. The purpose of this chapter is to assess the validity of several performance measures by examining the correlation between such measures and the program impacts discussed in Chapter 4. This chapter has two sections. The first section discusses job-entry and off-welfare measures. For all indicators, both unweighted and weighted versions are formulated. The unweighted ones count all enrollees equally; the weighted ones permit more weight to be given to program success in dealing with the less employable, more dependent subgroups. The second, and last, section addresses program participation and coverage indicators.

I. Job-Entry and Off-Welfare Measures

It is often argued that counting job entries and welfare case closures focuses the attention of program managers on the true program objectives, namely, increasing employment and decreasing welfare dependency. Emphasis on recording job entries can also serve an administrative function in AFDC eligibility determination, by encouraging and providing program staff with an incentive to make sure that client earnings are accurately reported to Income Maintenance Offices and that grant reductions are made where appropriate.

These outcomes are only valid indicators of underlying performance, as noted, if they are, in fact, correlated with real program impacts. If there is such a correlation, a high rate of job entry would indicate a large impact on employment, and a low rate would indicate a small impact, and similarly for AFDC case closures.

A. How Useful Are the Outcome Measures?

Several empirical observations already discussed strongly imply that the correlation of outcomes with impacts may not, in fact, be a strong one. For one thing, as the experience of the control groups in this analysis has shown, many recipients find jobs and leave welfare in the absence of program assistance. Most employment and welfare departures recorded for the experimental groups, in other words, would have occurred without the special intervention. Outcome measures of performance, therefore, clearly overstate true performance, and the overstatement is quite large. Thus, operators of welfare employment programs could record substantial numbers of "placements" and large "welfare savings" without providing any real services at all and without changing behavior in any way.

In this regard, it is telling that within the experimental groups examined for this study at least as much employment and as many welfare case closures accrued to nonparticipants as to those who actually participated in services.[1] This does not mean the programs had larger impacts on nonparticipants than participants, or even that they had any impacts on nonparticipants. It means only that substantial numbers were able to find jobs with minimal assistance and encouragement.

Simple outcome measures also cannot provide meaningful comparisons of performance across programs or components or from year to year. Some programs' high rates of job entry may result from their having a relatively "job-ready" target population or a strong labor market; the apparently poor rate of job entry for other programs may stem from a less job-ready target population or a poorer labor market. Because San Diego enrolled only applicants and Cook County enrolled current recipients and approved applicants, their rates of job entry were destined to be different from each other, even had they implemented identical programs. Follow-up employment rates for experimentals in each of the five programs studied differed substantially, but those differences did not reflect primarily differences in program impact. Most of the differences were determined by the subgroup composition of the enrollees, by local AFDC eligibility regulations, and by area labor market conditions. With the exception of Cook County, program impacts on employment were similar across programs. In other words, a lower post-program employment rate did not signify a lower program impact.

For these two reasons -- overstatement and misleading cross-program comparisons -- simple outcome measures do not fulfill the monitoring function required of a valid indicator of performance. Quite as serious as either of these problems is the fact that the degree of overstatement of performance differs substantially for different subgroups of program eligibles. In the preceding chapter, it was shown that individuals in the top employability categories typically experience post-program employment rates two or even three times as large as groups in the lower categories. The dramatic differences in job entry rates that this implies provide a strong incentive for program operators to pay the most attention to individuals at the top. Conscientious program administrators seeking high job-entry rates may focus staff time and resources on placing relatively job-ready registrants, many of whom might have been able to find jobs on their own. Administrators are given no incentive to test or implement services that would be effective for the least job-ready subgroups. This consideration can be even more important for voluntary programs than for mandatory ones, since outreach and screening have greater scope in determining the size and composition of the enrolled population when participation is not required.

An additional examination of the correlation between outcome measures and impacts was undertaken with the experimental Work/Welfare data. For this purpose, a short-term job entry was defined as employed at some point during quarters 2 or 3 after random assignment, and short-term off-welfare status was defined as receiving no welfare payments in the third quarter. Somewhat longer-term measures took into account quarter 4 and the following ones for employment, and quarter 6 for welfare payments. It should be noted that job entry rates derived from UI earnings will be higher than the actual job entry rates reported by these programs because the UI data provide a more complete record of employment than the contact information available to program staff.[2]

Table 5.1 displays, in summary form, the results of correlating the short-term job entry and off-welfare outcomes with program impacts on earnings and welfare payments estimated for each experimental group member on the basis of regression results for San Diego, Baltimore, Virginia, Arkansas, and Cook County.[3] The indicators are ranked in the table as follows:

TABLE 5.1

AFDC EXPERIMENTALS: VALIDITY OF SIMPLE JOB ENTRY AND OFF-WELFARE PERFORMANCE INDICATORS

	Correlation of indicator JOB ENTRY with impact on:	
Program	Earnings Gain	Welfare Savings
San Diego	poor	weak
Baltimore	weak	weak [fair]
Virginia	GOOD	poor
Arkansas	GOOD	weak
Cook County	GOOD	fair

	Correlation of indicator OFF WELFARE with impact on:	
Program	Earnings Gain	Welfare Savings
San Diego	fair	weak
Baltimore	fair	fair
Virginia	fair	poor
Arkansas	fair	fair
Cook County	GOOD	fair [GOOD]

SOURCE: See Table 4.1.

NOTES: This table summarizes the correlations between the designated indicator and earnings gains or welfare savings. The following symbols are used:

GOOD	correlation has the correct sign and is statistically significant
fair	correlation has the correct sign but is not statistically significant
weak	correlation has the wrong sign but is not statistically significant
poor	correlation has the wrong sign and is statistically significant

A longer-term version of the indicator was also tested in a second procedure by examining its partial correlation with the predicted impact while controlling for the short-term indicator. If the partial correlation of a longer-term version raised the indicator's rank from one of the two lower to one of the two higher ratings, or from "fair" to "GOOD," that change is noted in brackets in the table. "Short-term" and "longer-term" indicators are defined as follows:

Short-term job entry	Any UI earnings quarters 2 or 3
Longer-term job entry	Any UI earnings quarters 4 through last
Short-term off welfare	No AFDC payments quarter 3
Longer-term off welfare	No AFDC payments quarter 6

Rating	Correlation between indicator and impact
GOOD	positive, statistically significant
FAIR	positive, not statistically significant
WEAK	negative, not statistically significant
POOR	negative, statistically significant

Rankings are provided for all short-term versions of these indicators. If the longer-term version indicated substantial improvement, the higher rank is shown in brackets. It should be noted that a result of "good" means only that the indicator has some validity when applied to program enrollees as a group; it does not imply that the measure is a reliable indicator that the program has or has not had an effect for any individual enrollee.

The ideal indicator would have a good rating with both earnings gains and welfare reductions. The actual correlations fell short of this ideal. For two of the programs (San Diego and Baltimore), the job-entry measure was clearly not consistent with actual performance; short-term job entry was a weak or poor indicator of earning impacts, and the longer-term version showed little improvement. Job entry was also not a satisfactory indicator of welfare savings, nor was the off-welfare measure.

For the other three programs -- Virginia, Arkansas and Cook County -- job entry had good ratings for earnings impacts. However, only for Cook County was there any consistency between the correlation of job entry with both earnings and welfare reductions, and even there the ratings were not identical. For Virginia and Arkansas, job entry was a poor or weak indicator of welfare reductions. This is consistent with the subgroup findings: Virginia and Arkansas achieved their largest earnings impacts at or near the top of the dependency spectrum, whereas welfare reductions occured primarily among sample members further down the dependency spectrum, where job entry rates were lower.

The short-term off-welfare measure was slightly better. It had fair ratings as an indicator of earnings gains in four of the programs and a rating of good in Cook County. It had a poor rating as an indicator of welfare savings in Virginia, but reached the level of fair for three of the five programs (Arkansas, Cook County, and Baltimore).

Interestingly, longer-term data did not lead to much better results. In only 2 of the 20 correlations in Table 5.1 did the longer-term version of an indicator improve the correlation. It should be noted, however, that the "short-term" outcome measures tested here make use of follow-up which is already longer than that available to many program operators, who often have only the enrollee's status at date of termination from the program.

These empirical correlations are not the only test of indicator validity, but they do highlight some of the disadvantages of simple job-entry and off-welfare performance measures. They confirm that outcome indicators are not always strongly correlated with underlying program impacts. They also indicate that, even in cases where job entry was correlated with earnings impact, this performance measure produced weaker results for reduction of dependency, a problem which concurrent use of case closure standards seems unlikely to fully correct. Job-entry standards were most compatible with earnings gains accruing to individuals who would have been off AFDC soon anyway. For administrators who wish to affect individuals who ordinarily would stay in the welfare system longer, simple outcome standards would not appear to encourage optimal allocation of program resources.

B. Can Better Measures Be Developed?

Up to this point, the job-entry measures have given equal value to all WIN-mandatory clients, regardless of their work and welfare histories. The disadvantage of such unweighted measures is that they contain incentives for program operators to serve clients unequally. This built-in bias is not necessarily in the direction of maximum impact, nor does it always lead to coordinated earnings gains and welfare savings. To reduce these problems, it is appropriate to consider a different scoring strategy -- one that gives more weight to job entries for registrants with weaker previous work records or longer time on welfare. The rationale for weighting is to retain the best property of outcome measures -- their emphasis on employment and departure from welfare as program goals -- while overriding their undesirable allocation properties.

Weighting may be simple or complex. It may mean simply keeping separate track of job entry rates for applicants and recipients or for individuals with high and low prior earnings. Administrators may then set different job entry standards for each subgroup and announce a higher priority (i.e., greater weight) on achieving job entries for one over the other. Or weighting may be based on a complex formula, involving a long list of enrollee characteristics, with the weights derived from regression analysis of employment and welfare behavior. Yet another kind of weighting takes the form of "waiting" -- providing a low-cost job search component to all enrollees to begin with and reserving more intensive services for individuals who have still not become employed after several weeks or months.

To find out whether weighting might improve indicator validity, the correlations in the preceding section were run on weighted job-entry and off-welfare measures. These new performance measures were based on a number of different weighting schemes; but all gave greater weight to successful outcomes for sample members with low predicted future earnings or long predicted length of time on welfare. Some of the tested weighting schemes used complex, regression-based indices of predicted earnings and welfare receipt of the kind already discussed. The poor correlations in San Diego and Baltimore improved for some of these weighting schemes, providing some evidence that giving extra weight in setting performance standards to the job entries to less employable welfare recipients may improve the link between performance measure and impact.

The most complex weighting systems utilize a complete demographic profile for each enrollee and assign a weight for each characteristic. This maximum use of available information may seem desirable, since it provides the most detailed weighting scheme. This level of detail may well be suitable for evaluations at the state and national level -- where additional weights can be calculated for local labor market conditions and AFDC statutory grant levels. It has drawbacks as a tool for local operators and caseworkers, however. First, the extra data collection is costly, and the more data needed the higher the chance of error. Second, the weights themselves must be estimated with care so as not to over-emphasize a variable that actually has relatively little operational importance. Third, and perhaps most important, the complexity of the formula may obscure rather than clarify the operational priorities line staff need. Because of these disadvantages, it is important to see whether the sought-after improvement in outcome indicators can be achieved by simpler weighting alternatives.

An alternative approach uses information about only the two best predictors of future employability and dependency -- namely, prior employment and welfare experience. One such measure can be created for job entries based only on prior employment:[4]

$3,000 or more earnings in prior year: 1 point per job entry
$1-2,999 earnings in prior year: 2 points per job entry
Not employed in prior year: 4 points per job entry

Another measure applies the same weights to length of prior AFDC receipt:

Never had own AFDC case: 1 point per job entry
Had own AFDC case for two years or less: 2 points per job entry
Had own AFDC case more than two years: 4 points per job entry

These or similar measures might be applied directly in scoring performance of program staff or local program units. The same principle would be embodied more flexibly in a directive to grant service priority to the lower subgroups of each set, with proper administrative controls to assure compliance.

The correlations of the two weighted measures with impacts on earnings and welfare payments are summarized in Tables 5.2 and 5.3, along with the corresponding increases or decreases in validity relative to unweighted measures. Increases in validity were recorded for 14 out of the 36 short-term correlations, decreases for three. These weighted indicators thus represented a moderate improvement over unweighted measures. For earnings

TABLE 5.2

AFDC EXPERIMENTALS: VALIDITY AND IMPROVEMENT OF JOB ENTRY AND OFF-WELFARE PERFORMANCE INDICATORS WEIGHTED BY PRIOR EARNINGS

	Correlation and improvement[a] of indicator JOB ENTRY WEIGHTED by PRIOR EARNINGS[b] with impact on:	
Program	Earnings Gain	Welfare Savings
San Diego	GOOD +	GOOD +
Baltimore	GOOD +	fair +
Virginia	poor -	weak 0
Arkansas	GOOD 0	weak 0 [fair +]
Cook County	GOOD 0	GOOD +

	Correlation and improvement[a] of indicator OFF WELFARE WEIGHTED by PRIOR EARNINGS[b] with impact on:	
Program	Earnings Gain	Welfare Savings
San Diego	GOOD +	GOOD +
Baltimore	GOOD +	GOOD +
Virginia	poor -	fair +
Arkansas	fair 0	fair 0
Cook County	weak -	GOOD +

SOURCE: See Table 4.1.

NOTES: See Table 5.1.

[a] Improvement over the unweighted version of the indicator is shown as follows:

+	correlation increased from poor or weak to fair or good, or from fair to good
-	correlation decreased from fair or good to poor or weak, or from good to fair
0	no change or minimal change

[b] Weights were assigned to job entry and off-welfare scores on the basis of prior earnings as follows:

Prior-Year Earnings	Points
$3000 or More	1
$1-2999	2
None	4

TABLE 5.3

AFDC EXPERIMENTALS: VALIDITY AND IMPROVEMENT OF JOB ENTRY AND OFF-WELFARE PERFORMANCE INDICATORS WEIGHTED BY PRIOR WELFARE

	Correlation and improvement[a] of indicator JOB ENTRY WEIGHTED by PRIOR WELFARE[b] with impact on:	
Program	Earnings Gain	Welfare Savings
San Diego	weak 0	poor 0
Baltimore	poor 0	poor 0 [-]
Virginia	GOOD 0	weak 0 [GOOD +]
Arkansas	GOOD 0	weak 0
Cook County[c]	---	---

	Correlation and improvement[a] of indicator OFF WELFARE WEIGHTED by PRIOR WELFARE[b] with impact on:	
Program	Earnings Gain	Welfare Savings
San Diego	GOOD +	poor 0
Baltimore	fair 0	fair 0
Virginia	GOOD +	weak 0
Arkansas	GOOD +	fair 0
Cook County[c]	---	---

SOURCE: See Table 4.1.

NOTES: See Table 5.1.

[a] Improvement over the unweighted version of the indicator is shown as follows:

+	correlation increased from poor or weak to fair or good, or from fair to good
-	correlation decreased from fair or good to poor or weak, or from good to fair
0	no change or minimal change

[b] Weights were assigned to job entry and off-welfare scores on the basis of prior welfare history as follows:

Had Own AFDC Case	Points
Never	1
Two Years or Less	2
More Than Two Years	4

[c] Length of prior welfare information was not available for Cook County.

impacts, the weighted measures predominantly had a rating of good. Much of the improvement came for San Diego and Baltimore, particularly with the prior earnings weights. For Virginia, Arkansas, and Cook County, weighted indicators usually ranked as high as unweighted, but not higher, although they would still be preferable in that they would tend to counteract the adverse targeting properties of the unweighted measures.

Weighting also made a difference to the correlations with welfare savings. In particular, off-welfare measures weighted by prior earnings showed good or fair correlations with welfare savings in all five programs. In addition, in all but two cases weighted off-welfare measures had correlations with earnings gains that were good or fair. This suggests that AFDC receipt data, which is accessible to social services agencies, may usefully supplement or substitute for employment data, which is often not available and, if available, not complete.

Weighting improved the indicators more than did extending the follow-up data included in the measures. For the simple weighted indicators shown, longer-term information only occasionally increased the correlations and only once changed a rank from fair to good. And only for 2 of the 36 correlations did the addition of longer-term data lead to further improvement in indicator validity than the weighting.

To assess the consistency of job entry performance measures with program cost-effectiveness, further analysis was undertaken with the subgroup cost data available for San Diego and Baltimore. Estimates of total gain to government budgets were produced from estimates of welfare and Medicaid reductions and increases in taxes paid, minus program costs. Total gains for program enrollees were estimated by subtracting welfare and

Medicaid reductions and any increases in taxes from increases in earnings. Job entries with simple prior earnings weights were found to be positively correlated with both total gains for government budgets and total gains for program enrollees.

Overall, then, weighting job entries by prior earnings or prior welfare receipt appears to constitute an improvement over the unweighted measure. It is worthwhile also to speculate on extensions of the weighting principle studied here. Additional objective factors (such as the absence of a high school diploma) might be given weight, as well as other factors relevant in particular local circumstances and program goals. Weighting could also be applied to other outcome measures, including wage rates and job retention. The weighting concept could be used to economize on data costs by assigning higher priority to collecting employment information for enrollees with lower prior earnings or a longer welfare history. In this connection, where a subsampling strategy is employed for the collection of follow-up data it is probably a good idea to make sure that the more dependent and less employable are sampled in statistically adequate numbers.

How much accuracy is required in setting relative weights? The research conducted for this study suggests that selecting appropriate characteristics for weighting is probably more important than precision in the weights themselves. Weighting by prior earnings worked somewhat better than weighting by AFDC history, but both are undoubtedly key predictions of future experience for WIN-mandatory AFDC caseloads. It would be difficult to justify a weighting scheme that ignored both these variables.

The patterns of rankings vary relatively little when the weight values changed. For example, when the 1-2-4 prior earnings weights were replaced with weights of 1-2-8 and 1-4-4, similar, though not identical, rankings emerged. Broad program directives, such as giving placement priority to enrollees with more than two years of AFDC history, are consistent with the weighting concept tested here. And, questions of technical validity aside, differential weighting may be seen as a method of increasing the incentives for serving more dependent subgroups of the WIN-mandatory caseload on grounds of equitability as well as efficiency. For this purpose, "accuracy" of weighting has a value dimension as well as a technical one, and judgmental adjustments to weights may therefore be seen as legitimate. Weights derived from regression analysis of a nationally-representative sample sidestep the political process, which is seen as an advantage by some and a disadvantage by others.

Weighting, whether simple or regression-based, is probably not the final word on performance monitoring. Even with much refined weighting schemes, some problems with outcome indicators are likely to remain.

First, it is not clear that weighting or regression adjustment can fully solve the problem of comparisons across areas and time. Even with regression adjustment, local performance scores can vary in ways that do not reflect real performance.[5] Second, although the indicators tested often ranked fair or good, none of them reliably indicated which particular employed enrollees were the ones impacted by the program and which would have found jobs anyway. The weighting approaches tested can therefore serve only as rough guides to local operators in allocating program slots among enrollees. For low-cost components, encouraging broad targeting with weighted indicators may be adequate. But for high-cost components, it becomes much more important to know who precisely will succeed by participating -- and only by participating. For that purpose gross demographic characteristics alone may not provide sufficient guidance. Future research in this area may focus instead on more complex methods of making individual targeting decisions, comparing formal and informal techniques of assessment, and evaluating screening and filtering devices such as preliminary job search or other low-cost employment-oriented activities.

II. Participation and Coverage

Performance measures based on program participation have often been used as an alternative or complement to employment and welfare outcome measures. Compared to outcome measures, participation rates have both advantages and disadvantages. One clear advantage is that participation can be easily observed in the short term. Management control over the day-to-day operation of a program can therefore be readily achieved by monitoring participation. In programs intended to be mandatory, monitoring participation may also be undertaken to ensure compliance among enrollees and equitable treatment on the part of operational units. And biasing the program toward serving the most job ready is probably not as strong with participation standards as with simple job entry or other outcome standards.

The impact findings imply that administrators who track participation can achieve positive results by aiming for broad involvement over a wide range of client types. At a minimum, participation standards should discourage concentration only on the most job ready. Monitoring total participation may not by itself be sufficient to accomplish this. Monitoring participation separately for key subgroups provides necessary information about what kinds of enrollees receive services and what kinds remain inactive. Participation standards can then be established to adjust the priority given to individuals with, for example, poor work records or long welfare history. Priorities for service can be set explicitly. Or a weighting scheme like that just applied to job entries can be used to provide incentives in a less rigid fashion.

Some important questions about standards for participation are not answered by this study. On the operational level, it has often been observed that the actual nature and intensity of participation for components with the same name varies widely across localities. For example, the number of employer contacts required of registrants in independent job search is small in some areas, large in others. For group job search, there is variability in the content of sessions and the amount of time spent in attendance. Quality of participation may therefore be important to monitor, but not much can be said about this topic here.

The participation rates given in Chapter 2 may not be a suitable guide for planning. Those rates were calculated for research purposes. They are "ever participated" rates rather than the point-in-time participation rates usually available to program operators. In addition, comparisons across localities are not straightforward, just as with outcome indicators. Determining reasonable overall participation goals will hinge on characteristics of eligibles, local labor market conditions, program objectives, and available resources. These problems are taken up in greater depth in a forthcoming MDRC study.[6]

Participation monitoring faces one important conceptual issue not faced by outcome monitoring. The absence of a job entry clearly indicates the absence of a favorable employment impact for an individual. In contrast, nonparticipation does not necessarily mean that a program has failed to "reach" an individual as planned. In mandatory programs, sanctioning and other program contacts with nonparticipants are explicitly intended to handle noncompliance and to affect the behavior of enrollees. Some program impacts are therefore expected on nonparticipants. In addition, enrollees may find work or leave welfare in lieu of participating, and these responses are also part of program impact. For these individuals, the program objective is achieved without participation; having them participate would not only increase program costs but might also delay the employment or case closure outcome. Such contingencies are not accounted for under common operational definitions of "participation." Moreover, a drive for maximum participation may not be efficient. It may result in wasteful program expenditures on many enrollees who would have become employed or left welfare without participating.

To handle the difficulty, MDRC has used the concept of program *coverage*. Coverage measures have considerable potential, although to date they have been used only in evaluation research for the MDRC Work/Welfare demonstration. These measures count, in addition to instances of participation *per se*, cases with some acceptable substitute for participation or where sanctions for nonparticipation have been imposed. The concept of coverage takes into account normal welfare caseload turnover, but it does so without requiring information about prior employment and welfare and it need not involve weights.

Under a coverage formula, a client might be counted as "covered" by program requirements if any of these outcomes is achieved:

1. Participates in program activities
2. Becomes employed
3. Leaves AFDC
4. Is sanctioned for nonparticipation

The incentive effects of coverage standards are opposite to those of job entry or case closure standards. To maximize coverage, the attention of administrators is automatically directed to potential longer-term recipients. Individuals remaining on welfare only a short time will automatically be counted as covered when they leave AFDC; provision of services for them will not add to the program's coverage score because they will already be counted. On the other hand, those individuals who have longer expected length of stay on welfare can be covered only if they are reached by some program component. A coverage standard therefore carries a built-in incentive to work with individuals below the top tier.

The following example suggests how a coverage measure might work in practice. In the five programs studied, only 5 to 25 percent of experimentals were still on welfare nine months after enrollment and had not begun employment, had not participated in any major component, or had not been sanctioned for not participating. Thus, the 9-month coverage rate ranged from 75 to 95 percent of enrollees. These rates can convey a meaningful overall impression to legislators and the public about how well a program is reaching its eligible caseload. In addition, the goal of increasing coverage would shift attention toward more dependent subgroups because they are typically on welfare and enrolled in the program longest. In this connection, it is noteworthy that two-thirds of the "not-covered" experimentals were recipients, and three-quarters of this group had no prior earnings.[7]

There are, however, important disadvantages to the use of coverage indicators. Operationally, this class of measures presupposes the capability to follow the participation status of individual enrollees over time, which would require expenditures on setting up and operating tracking systems. Moreover, the cost of collecting, coordinating, and quality checking data on participation, welfare receipt status, and time from enrollment -- all of which go into the coverage statistic -- may substantially exceed the cost of maintaining simple participation counts. Since coverage rates are highly sensitive to normal welfare turnover rates and area labor market conditions, they would not solve the problem of comparisons across localities.

In conclusion, two limitations of this study should be noted. First, no mention has been made of intermediate outcomes, such as literacy and basic and specialized skills. Monitoring improvements in these is becoming an increasing preoccupation of program managers.[8] Second, the nature of the programs included in the study prevented investigation of targeting for more expensive education and training services.

This report does not put forward one ideal set of program performance standards. It recognizes strengths and weakness in alternative measures. At the same time, it has endeavored to evaluate some general principles. The most fundamental is that in welfare employment programs performance measures should take account of differences in the job readiness and welfare dependency of the individuals served. They should do so in a manner that counteracts the "commonsense" notion that the best program results come from the top tier of eligibles. For this purpose, weighted outcome and participation measures correct some defects in the incentive properties of common unweighted measures. Coverage measures also hold promise.

APPENDIX A

APPENDIX A

STATISTICAL CONSIDERATIONS IN SUBGROUP COMPARISONS

This appendix considers some special statistical issues which arise in the analysis of subgroup impacts. Its purpose is to lay out the justification for the conclusions presented in the report summary. To do so, it focuses on the estimates for low-, mid- and high-dependency rankings shown in Table 4.9. Some tests of impact differences across selected subgroups are discussed. Approximate tests of statistical significance are examined to account for multiple impact estimates. Similar reasoning could be applied to any of the subgroup impacts presented elsewhere.

When impact estimates are available for an entire population, statistical tests are unnecessary. Whatever the estimate is, whether it is large or small, it may be accepted as the true aggregate program effect with certainty, assuming that the estimate was produced in a valid fashion. But estimates based on samples, rather than on an entire population, contain an element of chance. To help rule out the chance element and increase confidence in any recommendation for service priority, certain statistical principles can be applied to impact estimates. Basically, the larger the estimated impact for a particular subgroup, and the more consistently it is found across samples, the more likely it is that the program model is generally effective for that subgroup. In addition, the larger the estimated difference between the impact for a particular subgroup and the impact for the other subgroups, and the more consistently this difference is found across samples, the greater the confidence program planners may have that granting priority in services to that subgroup will maximize the effect of the services in question.

Two kinds of statistical tests of impacts are relevant for this discussion. First, the basic experimental-control difference (the estimate of program impact) for a subgroup can be used to evaluate the hypothesis that the subgroup obtained no program effect. This is the usual t-test, applied to the magnitude of the difference between the estimated impact and zero impact. The same kind of t-test may also be applied to the difference between the estimated impact and any other fixed number, such as $100 for earnings. Second, the magnitude of two impacts for different subgroups can be compared. This kind of test is necessarily much less precise for any given sample because it involves a comparison of two estimates rather than a comparison of one estimate with a fixed number.

Because there is more than one set of subgroups, an additional complication arises. It is clear that if an unlimited number of subgroups can be examined, then sooner or later some will turn out to show a statistically significant effect, even if chance alone is operating. This problem occurs in all research involving multiple comparisons, and it requires that the usual statistical tests be qualified and made more stringent.

The statistical criteria considered here are chosen to test whether certain targeting strategies are likely to achieve the results they are intended to achieve. There are several reasons why program planners might adopt a strategy of targeting. One reason is that they wish to maximize program effect on employment or welfare receipt, given the limited resources available for services. This is the reason which is addressed in this report. Several alternative statistical criteria -- some strict, some loose -- may be adopted to decide whether a particular group merits a recommendation for priority targeting in order to maximize program impact. Passing any particular criterion means that the associated level of confidence in the conclusions has been attained.

Perhaps the strictest criterion would be to require that a subgroup show consistently and statistically significantly greater impacts than the balance of the sample before stating that one particular subgroup has larger impacts than the rest. A less strict criterion would be to require the subgroup to show consistent and statistically significant impacts compared to zero -- that is, that the experimental-control difference for that subgroup be generally statistically significant. A still less strict criterion would be to require the subgroup estimates to be larger than the balance of the sample, but without exceeding usual statistical standards. The least strict criterion would be for estimates to exceed zero but without passing any statistical test.

It is the view of the author that only subgroups passing the most stringent criterion deserve a strong recommendation for priority targeting. The consistency requirement is, however, quite difficult to pass even with samples of a thousand or more. Moreover, it is clear that when resources are limited, some targeting decision must be made, and this necessity may legitimately warrant accepting a lower level of confidence until additional research findings become available. On the other hand, it is the author's view that if only the least strict criteria are passed, then recommendations to target, even with resource constraints, will not be appropriate. It should be added, that even if a stringent criterion is passed, there may be other considerations which figure prominently in a targeting decision and which qualify or override conclusions based on the impact data alone.

The first targeting hypothesis, suggested by common practice, is that it is worthwhile to focus attention on the most employable and least dependent enrollees. Given the preceding discussion, this hypothesis should be accepted only if impact estimates for these subgroups either exceed those of other subgroups with some regularity or exceed some cutoff value or, at a minimum, exceed zero by a statistically significant amount in a more or less consistent fashion. It is clear from inspection of Table 4.9 and other tables that this hypothesis is untenable. Not only do subgroups of the first tier fail to exhibit impacts above average or above the balance of the sample, but their estimates are also generally below average and are sometimes the lowest in any subgroup comparison.

The second hypothesis, which derives from previous empirical analyses of the distribution of expected welfare tenure across welfare subgroups, is that the least job-ready or most dependent should be given priority for services. The estimates in this report do not strongly support this hypothesis. Earnings impacts for subgroups classified as relatively dependent were not the largest and did not generally exceed the mid-dependency subgroups. Welfare savings did appear relatively large, especially in comparison to earnings, for some dependent subgroups, but this pattern was not consistent. Thus, although the observed savings are in line with theory, the estimates fail to pass either of the more stringent criteria established above. The inconsistency in welfare savings across samples reduces confidence and increases the risk of offering low-cost programs primarily to one of the most dependent subgroups.

Earnings impacts for subgroups in the middle dependency tier come the closest of the major subgroups to passing a test for priority services. Applicant returnees, in particular, show the most consistent earnings impacts across samples and exceed the $100 cutoff in the four programs that had overall earnings gains. Table A.1 presents some statistical tests comparing impacts for second-tier subgroups with those for subgroups below and above them. The table shows the sign of the impact differential between subgroups and its level of statistical significance and also gives the probability value associated with this cross-subgroup t-test. For example, applicant returnees had impacts greater than the remaining applicants plus all recipients (i.e., the balance of sample) in the three samples where such a comparison was possible, and the differential was statistically significant in two (i.e., in Baltimore and Arkansas but not in Virginia). Yet even this subgroup does not pass the most stringent criterion. The balance-of-sample test fails in Virginia, and the test of returnees against first-time applicants in San Diego (the only possible test there) is not statistically significant. Moreover, the impact for this subgroup in Cook County would not likely be statistically significantly positive, even if the necessary information were available to identify the group there.

A less stringent criterion is that impacts be consistently positive for a subgroup. Applicant returnees come closest to passing this criterion. In four of the five programs -- and in all four programs with overall positive and statistically significant impacts -- the estimates of earnings impacts are positive and statistically significant. Allowance should, however, be made for multiple comparisons in deciding whether this string of impacts is not the product of chance. There are, in fact, 14 major

TABLE A.1

AFDC APPLICANTS AND RECIPIENTS: TESTS OF DIFFERENCES IN IMPACTS ON EARNINGS AND AFDC PAYMENTS BETWEEN SELECTED MAJOR SUBGROUPS

	Test of Difference Between Quarterly Earnings Impacts				
Subgroup Comparison	San Diego	Baltimore	Virginia	Arkansas	Cook County
Applicant Returnees VERSUS					
All Other Subgroups	---	+ * (.08)	+ (.44)	+ ** (.02)	---
Other Applicants	+ (.27)	+ (.64)	+ (.32)	+ ** (.04)	---
Recipients	---	+ * (.06)	+ (.54)	+ ** (.03)	---
Applicants with Less than $3000 Prior Year Earnings VERSUS					
All Other Subgroups	---	+ *** (.01)	- * (.10)	+ (.14)	- ** (.02)
Other Applicants	+ * (.06)	+ ** (.05)	- *** (.00)	+ (.40)	- (.15)
Recipients	---	+ ** (.01)	- (.29)	+ (.17)	- ** (.02)
Applicant Returnees with Less than $3000 Prior Year Earnings VERSUS					
All Other Subgroups	---	+ ** (.02)	- (.43)	+ ** (.03)	---
Other Applicants	+ (.53)	+ (.15)	- (.23)	+ * (.06)	---
Recipients	---	+ ** (.02)	- (.61)	+ ** (.04)	---

(continued)

TABLE A.1 (continued)

Subgroup Comparison	Test of Difference Between Quarterly AFDC Payments Impacts				
	San Diego	Baltimore	Virginia	Arkansas	Cook County
Applicant Returnees VERSUS					
All Other Subgroups	---	- (.41)	+ (.87)	+ (.13)	---
Other Applicants	- (.41)	- (.92)	+ (.83)	+ (.46)	---
Recipients	---	- (.35)	+ (.90)	+ * (.06)	---
Applicants with Less than $3000 Prior Year Earnings VERSUS					
All Other Subgroups	---	- (.20)	- (.53)	+ (.21)	- (.52)
Other Applicants	- (.35)	- (.41)	- (.32)	- (.59)	- (.31)
Recipients	---	- (.22)	- (.67)	+ (.11)	- (.62)
Applicant Returnees with Less than $3000 Prior Year Earnings VERSUS					
All Other Subgroups	---	- (.41)	- (.59)	+ (.20)	---
Other Applicants	- (.21)	- (.78)	- (.49)	+ (.59)	---
Recipients	---	- (.32)	- (.68)	+ * (.09)	---

SOURCE: See Table 4.1.

NOTES: The table shows the signs and statistical significance of the impact of the subgroup at top of each panel minus the impact for the subgroup below it. Quarterly impacts are an average of quarters four through last. Two-tailed t-tests were performed for each program sample from a regression on pooled applicant and recipient data, except in San Diego, where recipients were not enrolled. Numerical values in the parentheses are the probabilities associated with the t-values. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; *** = 1 percent.

subgroups and subgroup combinations considered: 5 prior earnings categories, 5 prior welfare categories, and 4 combinations of prior earnings and prior welfare (referring to Tables 4.2, 4.3 and 4.4). These groups are not independent, but assuming that they are permits a conservative, if approximate, accounting for the multiple comparisons.

The odds of finding four out of five sample estimates greater than zero at the observed levels are quite low, even with 14 independent trials.[1] Consistency of impacts for this subgroup across program samples is less certain. For one thing, the weak earnings results for Cook County applicants make it virtually certain that at least one of the five programs failed to achieve earnings gains for returnees. Disregarding the estimates for Cook County, which did not obtain earnings gains for the sample as a whole, the likelihood that one of the other four programs obtained no real earnings effect for returnees is 0.055,[2] which passes the conventional statistical test criterion of 0.100.

This test is by no means satisfactory, however, since it indicates only that working with this subgroup should produce earnings impacts, not that these impacts are likely to be relatively large. A more stringent test would be to apply the $100 cutoff to subgroups in the four programs that achieved overall earnings gains, still assuming 14 comparisons. The probability of finding the observed array of earnings impacts for applicant returnees in the first four programs under the assumption that impacts for all subgroups are $100 is under 1 percent.[3] A more difficult test to pass is created by averaging the earnings impacts across the four samples. Even accounting for multiple comparisons, the resulting average exceeds $80 at conventional levels of statistical significance, although not the higher

$100 cutoff.[4] Finally, the odds that impacts were below these cutoffs in none of the four samples are on the order of 50-50, implying that even this subgroup does not pass the most stringent test of consistency.

This analysis of the principal results from a purely statistical point of view leads to the conclusions stated in the report. They imply that any strategy of priority targeting focused on the most employable or least dependent is not supported by the data, whatever the statistical criterion. The same applies to strategies to exclusively target the least employable and most dependent. Furthermore, when it comes to the most stringent criteria, none of the subgroups examined emerges with certainty as the best targeting choice.

Nevertheless, with lower degrees of confidence the applicant returnee and other mid-dependency groups may be identified as suitable for priority in services when resource constraints require that a choice be made. Weighing the constellation of statistical evidence together with other considerations, this report concludes that a recommendation for exclusive targeting is not supported; that there is no certain best choice for exclusive targeting; but that when targeting is imposed by a scarcity of resources the mid-dependency subgroups are a suitable starting point for seeking earnings gains with the kinds of low-to-moderate cost programs under study. For welfare savings the evidence supporting any exclusive targeting scheme is weak. The more dependent subgroups are likely to assume increased importance in achieving welfare savings, although the results were, again, inconsistent across samples and suggest some risk in working with the third tier.

FOOTNOTES

CHAPTER 1

1. Friedlander and Long, 1987.

2. Results of additional benefit-cost analyses were carried out by MDRC, which are quoted as relevant to this report.

3. The use of the term "placement" is avoided in the rest of this report. The term was originally used by the employment service to denote referral of a client to a particular job opening by program staff. It is therefore inappropriate for programs that rely on a client's own job search efforts. In addition, placements, or self-reported employment, tend to understate employment and earnings because recipients sometimes do not report jobs to welfare staff, or leave the program before they find a job and hence are not obligated to report their employment.

   Similarly, the term "off-welfare" is used rather than "case closure" because it is more inclusive. It covers persons who apply for AFDC, enter a program, but then quickly leave the welfare system without having been approved for a grant (i.e., without ever having had a case opened).

   "Off-welfare" is not identical to the "welfare reduction" indicators in use. The former looks only at whether families are receiving any AFDC payment, and can be stated either as a numerical count or as a percentage. The various welfare reduction formulae in use subtract pre-program welfare grant levels for clients from their post-program welfare receipt to arrive at a dollar figure, either aggregate or per registrant. This study tested an off-welfare indicator rather than a welfare dollar reduction indicator because the pre-program data necessary to simulate the latter were lacking from the San Diego and Baltimore research data bases.

4. See Bane and Ellwood, 1983; Ellwood, 1986.

5. See Ellwood, 1986, p. xii.

6. See O'Neill et al., 1984, p.84.

7. The role of performance scores in the actual distribution of funds has been quite small. The bulk of federal WIN funds have been allocated to states according to number of WIN registrants. On the basis of budget appropriations during the

1970s, it has been determined that incentive rewards for performance based on this formula could amount to about one-third of all federal WIN moneys given to states. (See Office of Family Assistance, 1985, pp. 13-14.)

In practice, annual funding changes have been restricted in other ways. WIN regional coordinators have had discretionary powers, and incentive moneys could be allocated for local performance achievements not incorporated in the mathematical formula or on the basis of other considerations. As a result, only about 3 percent of funds distributed in a given year have reflected performance scores, although cumulative changes across the years could have amounted to more. (Office of Family Assistance, 1985, p. 21.)

Job retention has been a more important determinant of the program performance score in the discretionary part of the WIN Allocation Formula than job entry, although there is some evidence that the complexity of the formula kept this fact hidden from line operators (Mitchell, Chadwin, Nightingale, 1980, p. 287). The relative potential of each element of the formula to raise a state's overall performance score differed, depending on how high or low its score on each element might be. The complexity of the discretionary part of the formula was such that determining which elements had the greatest influence on scores would be very difficult without sophisticated analysis and simulation.

8. Participation is observed now, whereas outcomes may be observed only after some months and may require substantial effort in locating clients to ask about their employment status. Monitoring subgroup participation may be the most effective way of ensuring local compliance with an optimal targeting plan.

9. The problem of specifying optimal performance standards for independent local service providers for JTPA programs has been highlighted by the growing use of fixed-priced contracting. The language of JTPA has encouraged the use of fixed-priced contracting because all costs incurred can be allocated to "training," thus helping programs to comply with the 15 percent cap on administration costs. For a thorough discussion of the possibilities and problems in fixed-priced contracting see Wallace, 1985.

10. Goldman, et al., 1986, p.92. Indicators that make use of pre-program client measures are often referred to as change-based indicators, with simple outcomes designated as level indicators. The example given in the text of this chapter for San Diego would suggest that change-based indicators should

prove superior to simple outcomes as proxies for real program impact. In that case, the change from no pre-program employment to employment during the follow-up period was associated with the larger program-induced impact on employment. The weighted job entry rates tested in this paper are change-based indicators, since they award more performance points for the employment of clients who were not employed in the recent pre-program period.

The relevant literature on indicator validation is based on several analyses of CETA. Borus, 1978, found that job entry had very little power to indicate net impact for CETA. Gay and Borus, 1980, in a study of four pre-CETA programs, found change indicators to be somewhat superior, and rated simple job entry as one of the poorest measures. In contrast, Geraci and King, 1981, found evidence supporting job entry as the better measure, as did Geraci, 1984. Zornitsky et al., 1985, produced results favoring level indicators. The latter three studies also concluded that post-program follow-up added valuable information about employment at the point of termination.

These studies all suffer serious methodological problems from having been based on non-experimental impact estimates. The principal issue -- the value of level indicators versus change-based indicators -- is still the most pressing one to be resolved in performance monitoring. The issue is complicated by the possibility that the best class of indicators may be different for welfare women, adult men and youth. Adult men entering employment programs typically exhibit a temporary pre-program dip in earnings, making prior earnings problematic as a proxy for earnings capability. Youth often have short and erratic earnings histories, and a pre-program earnings baseline may therefore be meaningless for them.

11. See Goldman, 1981; Wolfhagen, 1983; MDRC, 1980.

CHAPTER 2

1. See Goldman et al., 1986; Friedlander et al., 1985b; Riccio et al., 1986; Friedlander et al., 1987; Friedlander et al., 1985a. (For a summary of the demonstration's findings thus far, see Gueron, 1987).

2. In this report, participation and sanctioning rates were calculated on somewhat different bases than in the published state reports. In this study, the base is always "all experimentals." In the state reports, the base of "all program registrants" was often used. Most experimentals did, however, register for the programs, and the differences

between the figures cited here and those published in the state reports are not large.

3. In San Diego, a second experimental group received job search only. The program and its evaluation were also carried out for AFDC-Us. Neither of these research groups is analyzed in this study.

4. Cook County, like San Diego, had two research groups, one to test job search plus work experience and the other to test job search alone. Only the former is included in this subgroup study. The Cook County WIN Demonstration program also worked with AFDC-Us, but this part of the caseload was excluded from the evaluation.

5. Sample sizes in this report differ slightly from those in the corresponding state reports. An attempt was made here to assign values to demographic data where these were missing. If missing data could not be inferred with reasonable certainty, the cases were dropped from the analysis. The effect of this strategy on sample size was the gain of 7 cases in San Diego, 54 cases in Baltimore, and 8 in Arkansas, but a loss of 32 cases in Virginia. The Chicago sample was unchanged.

   Randomization produced similar experimental and control groups with some differences. There were small differences between research groups in ethnicity and marital status in the San Diego sample. In Baltimore and Virginia, small differences were apparent in measures of education, prior employment and earnings.

6. This does not mean that the indicated subgroups account for the bulk of all AFDC expenditures. Benefits paid to families outside the WIN-mandatory sample are not counted. Nationally, about two-thirds of AFDC families are WIN-exempt.

7. In discussing the ability to predict differences in behavior, a distinction must be made between individuals and groups. As shown in the text, a wide range of differences in average outcomes across groups can be predicted quite well. Differences among individuals are less predictable.

8. Prior work and welfare histories are important in these adult samples, but may not play the same roles for a sample of younger mothers. Youth, simply because they are young, often have short work and welfare histories making these predictors less powerful than for adults.

9. Younger children are defined here as 12 years or under. Since, with the exception of Arkansas, these programs employed the traditional WIN definition of mandatory, women with

children under age six were largely exempt. Those few who are in the study samples are probably not representative of the rest. Restriction of the study sample to WIN-mandatories implies that correlations of demographic characteristics with future employment and welfare receipt may not be representative of the AFDC population at large, or of the wider population of poor family heads.

10. An inference based on the observed patterns of participation.

CHAPTER 3

1. For more complete reports of data quality control, see the individual state reports.

2. For more detail about data sources and follow-up, consult the state reports.

3. The distinction between unconditional and conditional impact estimates can be developed as follows. The basic impact regression model is

$$Y(T, S1, S2, X)$$

where

Y	outcome variable
T	experimental group dummy variable
S1	dummy variable for subgroup dimension 1
S2	dummy variable for subgroup dimension 2
X	vector of additonal control variables

The full sample impact is the coefficient of T. The unconditional subgroup estimates for S1 come from the regression model

$$Y(TS1, TNS1, S1, S2, X)$$

where

$$TS1 = T * S1$$

$$TNS1 = T * (1-S1)$$

The impact on groups S1=1 and S1=0 are read from the coefficients of TS1 and TNS1, respectively. Finally, the conditional model is

Y(T, TS1, TS2, S1, S2, X)

where

TS2 = T * S2

and the coefficient of T is the impact when S1=0 and S2=0. The coefficient of TS1 is the additional impact attributable to the S1 characterstic when S2 is held constant. The coefficient of TS2 is the additional impact attributable to the S2 characteristic when S1 is held constant.

Interactive specifications are possible for both unconditional and conditional models. For the unconditional case,

Y(TS12, TS1N2, TSN12, TSN1N2, S1, S2, S12, X)

where

TS12 = T * S1 * S2

TS1N2 = T * S1 * (1-S2)

TSN12 = T * (1-S1) * S2

TSN1N2 = T * (1-S1) * (1-S2)

S12 = S1 * S2

For the conditional case,

Y(T, TS1, TS2, TS12, S1, S2, S12, X)

Coefficients in this latter model can be combined to reproduce the unconditional interaction estimates exactly. But when a third subgroup dimension is introduced, S3, the term TS3 in the conditional model would make the two sets of interaction estimates different.

4. See Borus, 1978.

5. Individual impact estimates are made by (1) regressing demographic and background characteristics on employment and welfare outcomes for the experimental and control groups, and then (2) using the coefficients obtained from these regressions, along with the characteristics of individual members of the experimental group, to predict individual impacts. The first stage estimate is made from the conditional subgroup impact regression model. That is, from the regression that

contains the full array of experimental subgroup interactions, a prediction is made for the expected program impact on earnings and welfare receipt for each person in the experimental sample. The net impact estimate will differ for each person, depending on the demographic and prior work and welfare characteristics at the time of entry into the research sample.

These are sometimes referred to as direct estimates. For example, with treatment interactions for prior employment, education and number of children, one impact would be predicted for an experimental with no prior employment, no diploma, one child; a different net impact would be predicted for an experimental with any difference in any of these characteristics. The more variance in the dependent variable that can be accounted for by the regression model, the better the predicted net impacts. At the present state of knowledge, however, most of the variation in the outcome measures cannot be explained.

CHAPTER 4

1. For this analysis, earnings impact regressions were run on the pooled sample of applicants and recipients, separately in Baltimore, Virginia, Arkansas and Cook County. The model specified an experimental group dummy, a dummy for applicants, and a dummy for an experimental-applicant interaction. This last dummy gave the estimate of the unconditional impact differences. Interactions of experimental group membership with all other subgroup characteristics were then added and the same coefficient read again. The coefficient changed very little. The t-statistic for this coefficient gives the statistical significance of the conditional difference in impacts between applicants and recipients. Applicant/recipient differences in earnings gains were statistically significant in Baltimore and Cook County but not in Virginia and Arkansas.

2. This analysis of subgroup combinations intentionally does not break up the first-time applicant group into prior earnings subgroups. Continually subdividing the samples in this fashion quickly reduces sample sizes below the point where statistical analysis is meaningful. Moreover, it is questionable whether very small subgroups can have much policy importance for mass participation programs.

3. Friedlander, et al., 1985b.

4. Friedlander, 1987.

5. AFDC benefit levels also varied across counties in Virginia.

6. See Riccio et al., 1986, p. xiv.

7. Length of employment usually means job retention, i.e., remaining continuously employed with a particular employer. It is difficult to identify this kind of job retention with UI earnings data. It is, however, possible to examine other measures of attachment to employment. For example, it was found that 30.3 percent of applicant experimentals worked in six or more quarters from quarter 4 through the last follow-up quarter, a statistically significant increase of 8.5 percentage points over applicant controls. The corresponding level for recipient experimentals was 11.7 percent, and this represented an increase of only 2.7 percentage points over recipient controls (not statistically significant).

8. To determine the relationship between impact on employment and impact on welfare receipt for recipients, short- and longer-term employment variables were added to the right-hand side of the welfare impact regression. The resulting equation may be interpreted as part of a recursive model of program effects:

$$E(T, X)$$

$$A(E, T, X)$$

where the variables are defined as follows:

E = the set of follow-up employment and earnings variables

A = the set of follow-up welfare receipt variables

T = the treatment dummy

X = the set of control variables.

As elsewhere in this report, A was defined as welfare payments for quarter 4 through the last follow-up quarter. Several specifications for E were tested to determine how their introduction into the second equation would change the coefficient of T from the simple impact equation A(T, X). The largest change was to reduce welfare savings for Arkansas recipients from $60 per quarter down to $46, a decline of less than 25 percent. Thus, at least three-quarters of the impact on AFDC payments for this subgroup came through program effects other than an increase in UI earnings.

9. See Friedlander et al., 1986.

10. For the method of calculation of this percentage, see note 15.

In San Diego, the total dollar savings for unmarried women were 123 percent of the total for the full sample. For Virginia, this figure was 104 percent, and for Arkansas, 88 percent. The unweighted average of these three is 105 percent. This figure may be interpreted to mean that if these programs had served unmarried women in the same way they did, but served no one else, then their total welfare savings would have been slightly higher than they were. This assumes no interactions between subgroups. For never-married women, the shares were 39 percent for San Diego, 87 percent for Virginia, and 91 percent for Arkansas, giving a simple average of 72 percent.

11. This test of statistical significance is not relevant to the hypothesis that impacts are smaller for women with a pre-school child. There is a high degree of confidence that this hypothesis is not true for this sample simply because the employment and earnings effects are larger for this subgroup rather than smaller. The absence of statistical significance means only that it is not certain whether the impact for this subgroup would tend usually to be larger on repeated trials.

12. It was, moreover, not possible with the available data to identify any subgroup of applicant returnees, or certain of the other subgroups examined above. Readers of the final Cook County report (Friedlander et al., 1987) will recognize that "applicants" were then labeled "new recipients" and "recipients" were called "prior recipients."

13. Friedlander et al., 1987, pp. 80-83.

14. The odds of one or more of the subgroups attaining earnings gains statistically significant at the level indicated purely by chance are less than one in ten.

15. The total earnings impacts for a program were calculated by multiplying the average impact per experimental by the number of experimentals. The portion of the impact associated with applicant returnees was calculated by multiplying the per-experimental impact for this subgroup by the number of experimentals in the subgroup. Division of the latter amount by the former converts the share from a dollar figure into a proportion. These proportions were 89 percent for San Diego, 69 percent for Baltimore, 49 percent for Virginia, and 77 percent for Arkansas. The simple average of these is 71 percent, the top end of the range cited in the text. Excluding San Diego, which did not work with recipients, the average is 65 percent, the bottom end of the range cited. These figures should be interpreted to mean that if the programs had served applicant returnees exactly as they did and did not enroll any other subgroups, the total earnings

impact would be 65-70 percent as large as what was actually measured, assuming that effects for any subgroup are independent of effects for the others.

CHAPTER 5

1. Employment and welfare receipt rates were calculated for nonparticipants and participants in San Diego, Batlimore, and Virginia. No less than half of all job entries in each of these programs were obtained by experimental sample members who never became active in any formal program component. Nonparticipants outnumbered participants, and the percent of nonparticipants who held UI-recorded employment at quarter 6 was at worst only a few percentage points lower than for participants. In addition, nonparticipants were more likley to be off AFDC by that time.

2. Under-reporting of job entries can occur when case heads who leave welfare because they have found jobs do not report employment. Particularly in large urban areas with large caseloads, cases are often closed because the client fails to respond to some attempt at contact, making it impossible to record employment status or other eligibility factors. In addition, reports of employment obtained by income maintenance staff for the purpose of adjusting grant payments are not always reported back to the staff of the employment program.

3. Regressions for average earnings and average welfare payments over quarter 4 through the last quarter were run with all treatment-subgroup interactions in the model at once. The coefficients of these interactions were then used to predict for every experimental group member the expected net impact on earnings and welfare receipt. These new variables were then correlated with employment and off-welfare status, using only the experimental group sample.

4. These weights represent approximately the relationship of control group mean earnings for prior-earnings categories in the composite impact table in the impact chapter of the earlier report in this study.

5. Some of the variety in local job-entry rates cannot be accounted for by current regression models, and this problem becomes more severe the greater the degree of disaggregation. To examine this issue, the local office designators available in San Diego and Cook County were used. Regression-adjusted job-entry rates and average earnings impacts were calculated for each office, using the subgroup demographics as regression control variables. A crude estimate of "mistakes" made by regression-adjusted scores may be made by counting the number

of offices with high adjusted job-entry rates and low impact, or vice versa. In San Diego, there were 7 offices. Of the 3 offices with adjusted job-entry rates above the median, 2 had impacts below the median. Two of the 3 offices with job entries below the median had impacts above it. In Cook County, there were 11 offices. Two of the 5 offices with job entry rates above the median had impacts below the median, and 1 office with job entries below the median had impacts above it. These results should be considered suggestive rather than definitive. Future research may well succeed in identifying labor market variables suitable for the local office level of aggregation; this analysis did not utilize such variables.

6. Hamilton, 1988.

7. Based on statistics for San Diego, Baltimore, and Virginia. The data for Arkansas and Cook County were not available in a form that permitted estimates to be made.

8. Intermediate outcomes are program objectives intended to lead eventually to employment and departure from welfare. One major disadvantage in emphasizing intermediate outcomes, however, is that they do not necessarily bring about impacts on employment and welfare receipt. On the other hand, deficits in skills are more readily measured than deficits in "employability." Likewise, compared to the ultimate employment and welfare outcomes, improvements in these kinds of intermediate outcomes are less likely to occur without some special training. For example, fluctuations in local unemployment rates or changes in family circumstance will greatly affect the probability of a job entry, independent of anything a program might do. But such fluctuations have little influence on reading level, which is likely to change only with participation in some remedial activity. Thus, if a program obtains increases in reading level for a high percentage of enrollees, it may take credit for success. In theory, administrators may be more confident that this kind of success is a product of the program's efforts than they may be for job entry and off-welfare outcomes.

   In practice, this statement probably requires qualification. Enrollees even in mandatory welfare employment programs show a surprizingly high degree of self-initiated educational activity. For example, control-group members in Virginia, where data was obtained on non-program educational activity, showed rates of activity as high as experimentals, who could be referred to educational institutions as part of their program participation. Similar data were available for Cook County as well, and educational participation reached about 17 percent for experimentals and controls, even though the program itself did not actively refer clients. To the degree that a program provides educational services to an individual who would have found them anyway, the improvement in skills is not a program effect.

APPENDIX A

1. This probability is calculated in the following fashion. First, the right-tail probabilities associated with the t-statistics for applicant returnee earnings impacts for San Diego, Baltimore, Virginia and Arkansas are multiplied by each other. These probability values are, respectively, 0.0058, 0.0039, 0.0427 and 0.0032. For Cook County, the probability was assumed to be unity in the absence of actual data. The product is the probability of obtaining the observed impact values or higher, assuming that all the true values are zero. The probability of obtaining lower estimates in 14 independent trials is the 14th power of 1 minus this probability. This result is then subtracted from 1 to yield the probability referred to in the text: less than one in a million.

2. The likelihood that one of the four programs achieved zero or lower earnings impact for applicant returnees is 1 minus the likelihood that impacts were greater than zero in all four. The computation therefore begins by subtracting each of the four probabilities cited in the previous footnote from unity. The figure in the text is then 1 minus the product of these four numbers.

3. The figure in the text is calculated as in footnote 1, substituting the probability values associated with a t-test against $100 instead of zero. The four probability values are 0.1754 for San Diego, 0.1063 for Baltimore, 0.4147 for Virginia and 0.0744 for Arkansas. The result is 0.008.

4. Using equal weights for each sample, the average exceeds $100 at the 5 percent level using a standard one-tailed t-test. The same result holds if inverse standard errors of the impact estimates are used for weights in order to minimize the variance of the average. Allowing for 14 repetitions and maintaining an experiment-wise error rate of 10 percent with a one-tailed test, these averages are not statistically significant against the $100 cutoff, although they are against an $80 cutoff. This test is an extremely conservative approach to multiple comparisons and indicates an acceptable degree of confidence that earnings impacts for applicant returnees on average fall above at least the lower cutoff.

REFERENCES

Bane, Mary Jo; and Ellwood, David T. 1983. "The Dynamics of Dependence: The Routes to Self-Sufficiency." Cambridge, Mass.: Urban Systems Research and Engineering, Inc.

Borus, Michael E. October 1978. "Indicators of CETA Performance." *Industrial and Labor Relations Review*, Vol. 32, No. 1.

Ellwood, David T. January 1986. "Targeting 'Would-Be' Long-Term Recipients of AFDC." Princeton, N.J.: Mathematica Policy Research, Inc.

Friedlander, Daniel. 1987. *Supplemental Report on the Baltimore Options Program*. New York: Manpower Demonstration Research Corporation.

Friedlander, Daniel; Erickson, Marjorie; Hamilton, Gayle; and Knox, Virginia. 1986. *West Virginia: Final Report on the Community Work Experience Demonstration*. New York: Manpower Demosntration Research Corporation.

Friedlander, Daniel; Freedman, Stephen; Hamilton, Gayle; and Quint, Janet. 1987. *Illinois: Final Report on Job Search and Work Experience in Cook County*. New York: Manpower Demonstration Research Corporation.

Friedlander, Daniel; Hoerz, Gregory; Long, David; and Quint, Janet. 1985b. *Maryland: Final Report on the Employment Initiatives Evaluation*. New York: Manpower Demonstration Research Corporation.

Friedlander, Daniel; Hoerz, Gregory; Quint, Janet; and Riccio, James. 1985a. *Arkansas: Final Report on the WORK Program in Two Counties*. New York: Manpower Demonstration Research Corporation.

Friedlander, Daniel; and Long, David. 1987. *A Study of Performance Measures and Subgroup Impacts in Three Welfare Employment Programs*. New York: Manpower Demonstation Research Corporation.

Gay, Robert S.; and Borus, Michael E. Winter 1980. "Validating Performance Indicators for Employment and Training Programs." *The Journal of Human Resources*, Vol. 15, No. 1.

Geraci, Vincent J. August 1984. "Short-Term Indicators of Job Training Program Effects on Long-Term Participant Earnings." University of Texas-Austin, Center for the Study of Human Resources, Project Working Paper 2.

Geraci, Vincent J.; and King, Christopher T. September 1981. "Employment and Training (CETA) Program Performance: Long-Term Earnings Effects and Short-Term Indicators." University of Texas-Austin, Center for the Study of Human Resources, Working Paper Series Number 40-81.

Goldman, Barbara S. 1981. Impacts of the Immediate Job Search Assistance Experiment: Louisville WIN Research Laboratory Project. New York: Manpower Demonstration Research Corporation.

Goldman, Barbara; Friedlander, Daniel; and Long, David. 1986. California: Final Report on the San Diego Job Search and Work Experience Demonstration. New York: Manpower Demonstration Research Corporation.

Grossman, Jean Baldwin; Maynard, Rebecca; and Roberts, Judith. October 1985. Reanalysis of the Effects of Selected Employment and Training Programs for Welfare Recipients. Princeton, N.J.: Mathematica Policy Research, Inc.

Grossman, Jean Baldwin; and Mirsky, Audrey. February 1985. A Survey of Recent Programs Designed to Reduce Long-Term Welfare Dependency. Princeton, N.J.: Mathematica Policy Research, Inc.

Gueron, Judith. 1987. Reforming Welfare with Work. New York: The Ford Foundation.

Hamilton, Gayle. 1988. Interim Report on the Saturation Work Initiative Model in San Diego. New York: Manpower Demonstration Research Corporation.

Manpower Demonstration Research Corporation. 1980. Summary and Findings of the National Supported Work Demonstration. Cambridge, Mass.: Ballinger Publishing Company.

Mitchell, John J.; Chadwin, Mark L.; and Nightingale, Demetra S. 1980. Implementing Welfare-Employment Programs: An Institutional Analysis of the Work Incentive (WIN) Program. Washington, D.C.: U.S. Department of Labor, Employment and Training Administration, Monograph No. 78.

Office of Family Assistance, Department of Health and Human Services. April 1985. "A Study of the Work Incentives (WIN) Allocation Formula." Washington, D.C.: Office of Family Assistance.

O'Neill, June A.; Wolf, Douglas A.; Bassi, Lauri J.; and Hannan, Michael T. June 1984. An Analysis of Time on Welfare. Washington, D.C.: The Urban Institute.

Riccio, James; Cave, George; Freedman, Stephen; and Price, Marilyn. 1986. Virginia: Final Report on the Employment Services Program. New York: Manpower Demonstration Research Corporation.

Wallace, John W. January 1985. "Information Collection for the Job Training Partnership Act: Performance Standards vs. Evaluation." Washington, D.C.: National Commission for Employment Policy.

Wolfhagen, Carl. 1983. Job Search Strategies: Lessons from the Louisville

WIN Laboratory. New York: Manpower Demonstration Research Corp.

Zornitsky, J.; Schneider, G.; Sharick, R.; and Shapiro, E. June 1985. Post-Program Performance Measures for Adult Programs Funded Under Titles IIA and III of JTPA: Early Findings Prepared for the U.S. Department of Labor's Performance Standards Advisory Committee. Cambridge, Mass.: Abt Associates, Inc.

PUBLISHED AND FORTHCOMING STUDIES IN THE MDRC DEMONSTRATION OF STATE WORK/WELFARE INITIATIVES

MONOGRAPHS

Gueron, Judith. 1986. Work Initiatives for Welfare Recipients: Lessons From a Multi-State Experiment.

Gueron, Judith. 1987. Reforming Welfare with Work. (Published by the Ford Foundation.)

REPORTS

ARIZONA

Sherwood, Kay. 1984. Management Lessons From the Arizona WIN Demonstration Program.

ARKANSAS

Quint, Janet; with Goldman, Barbara; and Gueron, Judith. 1984. Interim Findings from the Arkansas WIN Demonstration Program.

Friedlander, Daniel; Hoerz, Gregory; Quint, Janet; Riccio, James; with Goldman, Barbara; Gueron, Judith; and Long, David. 1985. Arkansas: Final Report on the WORK Program in Two Counties.

CALIFORNIA

Goldman, Barbara; Gueron, Judith; Ball, Joseph; Price, Marilyn; with Friedlander, Daniel; and Hamilton, Gayle. 1984. Preliminary Findings from the San Diego Job Search and Work Experience Demonstration.

Goldman, Barbara; Friedlander, Daniel; Gueron Judith; Long, David; with Hamilton, Gayle; and Hoerz, Gregory. 1985. Findings from the San Diego Job Search and Work Experience Demonstration.

Goldman, Barbara; Friedlander, Daniel; Long, David; with Erickson, Marjorie; and Gueron, Judith. 1986. Final Report on the San Diego Job Search and Work Experience Demonstration.

ILLINOIS

Manpower Demonstration Research Corporation. 1985. Baseline Paper on the Evaluation of the WIN Demonstration Program in Cook County, Illinois.

Quint, Janet; Guy, Cynthia; with Hoerz, Gregory; Hamilton, Gayle; Ball, Joseph; Goldman, Barbara; and Gueron, Judith. 1986. Interim Findings from the Illinois WIN Demonstration Program in Cook County.

Friedlander, Daniel; Freedman, Stephen; Hamilton, Gayle; Quint, Janet; with Goldman, Barbara; Long, David; and Riccio, James. 1987. Final Report on Job Search and Work Experience in Cook County.

MAINE

Auspos, Patricia; with Ball, Joseph; Goldman, Barbara; and Gueron, Judith. 1985. Maine: Interim Findings from a Grant Diversion Program.

Auspos, Patricia; Cave, George; Long, David; with Hanson, Karla; Caspar, Emma; Friedlander, Daniel; and Goldman, Barbara. 1988. Final Report on the Training Opportunities in the Private Sector Program.

MARYLAND

Quint, Janet; with Ball, Joseph; Goldman, Barbara; Gueron, Judith; and Hamilton, Gayle. 1984. Interim Findings from the Maryland Employment Initiatives Programs.

Friedlander, Daniel; Hoerz, Gregory; Long, David; Quint, Janet; with Goldman, Barbara; and Gueron Judith. 1985. Maryland: Final Report on the Employment Initiatives Evaluation.

Friedlander, Daniel. 1987. Supplemental Report on the Baltimore Options Program.

NEW JERSEY

Final Report, 1988

VIRGINIA

Price, Marilyn; with Ball, Joseph; Goldman, Barbara; Gruber, David; Gueron, Judith; and Hamilton, Gayle. 1985. Interim Findings from the Virginia Employment Services Program.

Riccio, James; Cave, George; Freedman, Stephen; Price, Marilyn; with Friedlander, Daniel; Goldman, Barbara; Gueron, Judith; and Long, David. 1986. Final Report on the Virginia Employment Services Program.

WEST VIRGINIA

Ball, Joseph; with Hamilton, Gayle; Hoerz, Gregory; Goldman, Barbara; and Gueron, Judith. 1984. *West Virginia: Interim Findings on the Community Work Experience Demonstrations*.

Friedlander, Daniel; Erickson, Marjorie; Hamilton, Gayle; Knox, Virginia; with Goldman, Barbara; Gueron, Judith; and Long, David. 1986. *West Virginia: Final Report on the Community Work Experience Demonstrations*.

WELFARE GRANT DIVERSION

Bangser, Michael; Healy, James; and Ivry, Robert. 1985. *Welfare Grant Diversion: Early Observations from Programs in Six States*.

Bangser, Michael; Healy, James; and Ivry, Robert. 1986. *Welfare Grant Diversion: Lessons and Prospects*.

BROCHURE

Manpower Demonstration Research Corporation. 1987. "Findings on State Welfare Employment Programs."

OTHER

Goldman, Barbara; Cavin, Edward; Erickson, Marjorie; Hamilton, Gayle; Hasselbring, Darlene; and Reynolds, Sandra. 1985. *Relationship Between Earnings and Welfa]re Benefits for Working Recipients: Four Area Case Studies*.